

Universidad Autónoma de Madrid

Escuela Politécnica Superior



**Máster Universitario en Ingeniería Informática y Máster Universitario en
Investigación e Innovación en Tecnologías de la Información y las
Comunicaciones**

TRABAJO DE FIN DE MÁSTER

**MÉTODOS DE PREDICCIÓN APLICADOS AL PRECIO ELÉCTRICO
DEL MERCADO DIARIO**

Álvaro Romero Miralles
Tutor: Julia Díaz y José Ramón Dorronsoro

13/09/2017

MÉTODOS DE PREDICCIÓN APLICADOS AL PRECIO ELÉCTRICO DEL MERCADO DIARIO

Autor: Álvaro Romero Miralles
Tutor: Julia Díaz y José Ramón Dorronsoro

Departamento de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid

13/09/2017

Agradecimientos

En primer lugar, agradecer a José Ramón Dorronsoro y a Julia Díaz su dedicación, apoyo y comprensión durante el desarrollo de este trabajo, así como en los ámbitos académico y profesional.

También al resto de profesores tanto de la Universidad Autónoma de Madrid como de los estudios secundarios en el Instituto Salvador Dalí que supieron transmitirme la ilusión por las matemáticas, la informática y sus aplicaciones prácticas.

Igualmente, agradecer a mis compañeros del Instituto de Ingeniería del Conocimiento la facilidad con la que se trabaja junto a ellos y por dejarme ocupar los servidores con todas las ejecuciones para este trabajo con mucha paciencia.

En lo personal, gracias a mi familia, en especial a mis padres por su compromiso en mi formación. Y a la otra familia, la que se elige, a mis amigos puesto que sin su apoyo todo hubiera sido mucho más complicado.

Quiero agradecer aún más si cabe a mi pareja, que es quién quizá más haya sufrido el máster y este trabajo, incluso siendo capaz de aguantarme durante el verano con el portátil y el trabajo auestas.

Abstract

Abstract — During the last years the electrical systems around the world and in particular the Spanish electric sector have undergone great changes with the focus of turning them into more liberalized and competitive markets. For this reason, in many countries like Spain have appeared electric markets where producers sell and electricity retailers buy the power we consume.

All agents involved in this market need predictions of generation, demand and especially prices to be able to participate in them in a more efficient way, obtaining a greater profit.

The present work prospects the development of a tool that allows to predict the price of electricity for the next day in the most precise way possible.

For such target, this document analyzes the electric market to understand how prices are calculated and who are the agents that can make prices vary.

Traditional proposals in the literature range from the use of Game Theory to the use of Machine Learning, Time Series Analysis or Simulation Models.

In this project we have proposed a normalization of the target variable due to a strong seasonal component in an hourly and daily way to later benchmark several models of Machine Learning: Ridge Regression, K-Nearest Neighbors, Support Vector Machines, Neural Networks and Random Forest.

After observing that the best model is Random Forest, a discussion has been carried out on the appropriateness of the normalization for this algorithm. From this analysis it is obtained that the model that gives the best results has been Random Forest without applying the normalization function. This is due to the loss of the close relationship between the objective variable and the electric demand. The best Average Absolute Error obtained is 3.92€ for the whole period of 2016.

Finally, we propose an exploitation model based on a Software as a Service and an implementation design, using a kernel for the predictive model developed in Python and a web application developed in R with its library Shiny and everything encapsulated in a Docker container to facilitate its deployment.

Key words — Machine Learning, Big Data, Electric market, Predictive analysis, Prices, Random Forest

Resumen

Resumen — En los últimos años los sistemas eléctricos de todo el mundo y en concreto el español han sufrido grandes cambios con la directriz de convertirlos en mercados libres más competitivos. Por esta razón, en numerosos países como España han aparecido mercados eléctricos donde se compra y vende la energía que consumimos.

Todos los agentes que intervienen en este mercado necesitan de predicciones de generación, demanda y sobre todo precios para poder participar en ellos de una manera más eficiente, obteniendo unos mayores beneficios.

El presente trabajo es una prospección para el desarrollo de una herramienta que permita predecir el precio de la electricidad para el próximo día de la manera más precisa posible.

Para tal fin, se analiza el mercado eléctrico para entender cómo se calculan los precios y quiénes son los agentes que pueden hacer que varíe el precio.

Las propuestas que tradicionalmente se tratan en la literatura van desde el uso de Teoría de Juegos hasta el uso de técnicas de Aprendizaje Automático pasando por Análisis de Series Temporales o Modelos de Simulación.

En este proyecto se ha propuesto una normalización de la variable objetivo debido a su fuerte componente estacional de manera horaria y diaria para posteriormente realizar un *benchmark* de distintos modelos de Aprendizaje Automático: Regresión Ridge, K-Vecinos Próximos, Máquinas de Vectores Soporte, Redes Neuronales y *Random Forest*.

Tras observar que el mejor modelo es *Random Forest*, se ha llevado a cabo una discusión sobre la idoneidad de la normalización en este algoritmo. De este análisis se obtiene que el modelo que mejores resultados ha aportado es *Random Forest* sin aplicar la normalización debido a que se pierde la estrecha relación existente entre la variable objetivo y la demanda eléctrica, obteniendo un Error Absoluto Medio de 3.92€ para todo el periodo de 2016.

Por último, se ha propuesto un modelo de explotación basado en un *Software as a Service* y un diseño en el que se usan un núcleo para el modelo predictivo desarrollado en Python, una aplicación web desarrollada en R con la librería Shiny y todo ello encapsulado en un contenedor Docker para facilitar su posterior despliegue.

Palabras clave — Aprendizaje automático, Big Data, Mercado eléctrico, Análisis predictivo, Precios, Random Forest

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Alcance	2
1.3. Estructura del documento	2
2. Sistema eléctrico	5
2.1. Organización y características del sistema eléctrico español	5
2.1.1. Mercados <i>spot</i>	7
2.2. Cálculo del precio	10
2.3. Análisis de la serie temporal de precios	14
3. Estado del Arte	21
3.1. Modelos de Series Temporales	21
3.1.1. Características de las Series Temporales	22
3.1.2. Modelos autorregresivos	24
3.1.3. Modelo de media móvil	24
3.1.4. Modelos ARMA	24
3.1.5. Modelos ARIMA	25
3.1.6. Modelos SARIMA	25
3.1.7. Usos de modelos de series temporales para predicción de precios en la literatura	25
3.2. Modelos de aprendizaje automático	26
3.2.1. Regresión Lineal	26
3.2.2. Regresión Ridge	28
3.2.3. Árboles de decisión	28
3.2.4. K-Vecinos Próximos	30
3.2.5. Máquinas de Vectores Soporte	30
3.2.6. Redes Neuronales Artificiales	32
3.2.7. <i>Ensemble Learning</i>	35
3.3. Variables relevantes	36
4. Propuesta de servicio de predicción	39
4.1. Datos para los experimentos	40
4.2. Análisis de datos y normalización	40
4.2.1. Datos estadísticos	40
4.2.2. Detección de <i>outliers</i>	41
4.2.3. Generación de variables	41
4.2.4. Análisis de correlaciones	45

4.2.5. Normalización de la variable objetivo	46
4.3. Elección del modelo	49
4.3.1. Regresión Ridge	51
4.3.2. K-Vecinos Próximos (K-NN)	51
4.3.3. Perceptrón Multicapa (MLP)	51
4.3.4. Support Vector Machine (SVM)	51
4.3.5. <i>Random Forest</i>	52
4.4. Hiperparametrización	52
4.5. Métricas de error	53
4.6. Resultados	55
4.6.1. Análisis detallado del mejor modelo	56
4.6.2. Resultados de Test de Random Forest sin normalización	64
4.7. Modelo de explotación	64
5. Conclusiones y trabajo futuro	67
5.1. Conclusiones	67
5.2. Trabajo futuro	68
Apéndices	77
A. Tablas de resultados	79

Índice de tablas

4.1.	Tabla de estadísticas de los valores.	41
4.2.	Número de <i>outliers</i> por variable.	41
4.3.	Tabla de desviaciones típicas según la normalización aplicada	50
4.4.	Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de la Regresión Ridge.	55
4.5.	Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de K-NN.	55
4.6.	Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de MLP.	55
4.7.	Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de SVR	55
4.8.	Tabla del Error Absoluto Medio por mes y anual.	57
4.9.	Tabla del Error Absoluto Medio por mes y anual para demanda y precio normalizados.	58
4.10.	Tabla del Error Absoluto Medio por mes y anual para el mejor modelo: <i>Random Forest</i> con datos sin normalizar	64
4.11.	Tabla del Error Absoluto Medio por mes y medio para el mejor modelo en test: <i>Random Forest</i> con datos sin normalizar	64
5.1.	Resumen de resultados con distintas aproximaciones.	68
A.1.	Resultados hiperparametrización de Regresión Lineal Ridge.	81
A.2.	Resultados hiperparametrización de K-Vecinos Próximos.	83
A.3.	Resultados hiperparametrización de Random Forest.	86
A.4.	Resultados hiperparametrización de Perceptrón Multicapa.	87
A.5.	Resultados hiperparametrización de los modelos SVM.	89

Índice de figuras

2.1. Funcionamiento del sistema eléctrico español. <i>Elaboración propia.</i>	7
2.2. Horarios de los mercados de energía eléctrica. <i>Elaboración propia.</i>	8
2.3. Curvas agregadas de compra/venta y precio de la energía para el 22/08/2016 a las 00:00 <i>Elaboración propia.</i>	11
2.4. Oferta con condiciones complejas no cumplida. <i>Elaboración propia.</i>	12
2.5. Efecto de la exportación en la curva de oferta de compra. <i>Elaboración propia.</i> . .	13
2.6. Efecto de la importación en la curva de ofertas de venta. <i>Elaboración propia.</i> . .	13
2.7. Intercambio de energía con los países limítrofes a España. <i>Fuente REE.</i>	14
2.8. Curvas agregadas de compra/venta y precio de la energía tras exportaciones e importaciones y tras condiciones complejas para el 22/08/2016. <i>Elaboración propia.</i>	15
2.9. Histograma de precios diferenciado por día de la semana. <i>Elaboración propia.</i> . .	17
2.10. Media horaria por día de la semana. <i>Elaboración propia.</i>	17
2.11. Media horaria por día de la semana y mes. <i>Elaboración propia.</i>	18
2.12. Comparativa del precio medio horario en festivos contra no festivos. <i>Elaboración propia.</i>	19
3.1. Esquema de los modelos de predicción para precios de energía. <i>Fuente [48]</i>	22
4.1. Boxplot de las distintas variables. <i>Elaboración propia.</i>	42
4.2. Efecto de la temperatura T2 respecto de la original T. <i>Elaboración propia.</i>	43
4.3. Efecto de la temperatura T3 respecto de la original T. <i>Elaboración propia.</i>	43
4.4. Efecto de la temperatura T4 respecto de la original T. <i>Elaboración propia.</i>	44
4.5. Efecto de la temperatura T5 respecto de la original T. <i>Elaboración propia.</i>	44
4.6. Efecto de la temperatura T6 respecto de la original T. <i>Elaboración propia.</i>	45
4.7. Efecto de la temperatura T7 respecto de la original T. <i>Elaboración propia.</i>	45
4.8. Correlaciones de las distintas variables que afectan al precio. <i>Elaboración propia.</i>	46
4.9. Correlaciones de las distintas temperaturas definidas. <i>Elaboración propia.</i>	47
4.10. Proceso de validación realizado. <i>Elaboración propia.</i>	53
4.11. Proceso de validación cruzada tradicional. <i>Elaboración propia.</i>	53
4.12. Concepto de las métricas de error en regresión. <i>Fuente [8]</i>	54
4.13. Importancia de los distintos atributos. <i>Elaboración propia.</i>	57
4.14. Comparativa entre la demanda y el precio normalizado. <i>Elaboración propia.</i> . . .	58
4.15. Importancia de los distintos atributos normalizando la demanda igual que los precios. <i>Elaboración propia.</i>	58
4.16. Comparativa entre la demanda y el precio, ambas variables normalizadas. <i>Elaboración propia.</i>	59
4.17. Comparativa entre la demanda y el precio, ambas variables sin normalizar. <i>Elaboración propia.</i>	59

4.18. Importancia de las variables cuando no se normalizan. <i>Elaboración propia.</i>	60
4.19. Errores por hora y mes. <i>Elaboración propia.</i>	61
4.20. Resumen de errores por hora y mes para el año completo. <i>Elaboración propia.</i> . .	62
4.21. BoxPlot de errores con signo. <i>Elaboración propia.</i>	62
4.22. Histogramas de errores con signo. <i>Elaboración propia.</i>	63
4.23. Diseño del desarrollo de la aplicación. <i>Elaboración propia.</i>	65

1

Introducción

“The average introduction to almost any book is somewhat of a bore.”

Boris Karloff, *And the Darkness Falls*

1.1. Motivación

El sector eléctrico español y en concreto el mercado eléctrico es tremendamente complejo pero a la vez fundamental para poder mantener el modo de vida contemporáneo. En él es donde se compra la energía que llega a nuestros hogares e industrias y donde se vende la producción eléctrica generada en nuestras centrales.

En el mercado eléctrico existen una gran cantidad de agentes que negocian la energía y que establecen posiciones en función de la cantidad de energía que vaya a ser demandada y producida, de estrategias de sus propias compañías pero también teniendo muy en cuenta el precio al que se vaya a comprar y vender la energía. Por ejemplo, las centrales hidroeléctricas en las que hay más de una presa suelen tener motores con los que subir el agua a las que se encuentran en niveles superiores. De esta manera, estas centrales liberarán agua para producir energía durante las horas caras del día y demandarán energía para poner en marcha sus motores durante las horas de precios bajos de energía. Otro ejemplo pueden ser las paradas programadas para mantenimientos que realizan los parques eólicos o las plantas solares. Estas paradas actualmente se realizan teniendo en cuenta la cantidad de recurso eólico o solar que habrá durante el período de mantenimiento. Sin embargo, tener en cuenta el precio puede minimizar las pérdidas que supone realizar mantenimientos para las empresas generadoras.

Incluso en un futuro cuando las baterías sean una realidad, será necesario tener en cuenta el precio de la energía tanto para cargar las baterías, como para verter energía a la red en los momentos más idóneos.

En todas las situaciones anteriores, los involucrados en ellas se verían beneficiados si conociesen el precio de la electricidad pero, por desgracia, este precio es algo muy variable y que depende de muchos factores. La Ciencia de Datos apoyada en la informática y las matemáticas se convierte en una herramienta poderosa para crear una solución que permita conocer el precio de la electricidad en las próximas horas.

Este servicio de vital importancia mejoraría notablemente los procesos, minimizando pérdidas y maximizando ganancias a través del aporte de un conocimiento extra a los distintos agentes que deben tomar decisiones en cada una de las posibles situaciones que dependan del precio de la electricidad.

1.2. Alcance

El principal objetivo es la prospección tanto desde el punto de vista de la investigación como desde un punto de vista más práctico y dirigido a resolver un problema del «mundo real» como es conseguir un sistema de predicción de precios.

Dentro de los pasos previos al desarrollo del modelo como tal en este trabajo, al igual que en la gran parte de los proyectos de Ciencia de Datos, existen una serie de objetivos parciales como son por un lado, una revisión completa del funcionamiento del sistema eléctrico y, por otro lado, una revisión bibliográfica de los aspectos teóricos relevantes más usados durante la construcción del modelo, ambas con el fin de explicar de forma específica el problema y de reutilizar el conocimiento previo sobre la materia.

En este Trabajo de Fin de Máster se pone el foco en la definición del problema de manera correcta como un problema de aprendizaje automático que incluye, entre otras tareas, realizar un análisis descriptivo de las variables, definir métricas de error, encontrar un modelo o modelos que satisfagan la predicción de precios con el menor error posible, estudiar los resultados de cada uno de los modelos y hacer uso de esta información para de nuevo realizar una búsqueda de los mejores modelos. Más en concreto, se ha propuesto una función de normalización que reduzca la varianza de los datos y se han aplicado una batería de modelos al problema. Se ha encontrado que el modelo que mejor se adapta a la estructura subyacente de los datos para este problema ha sido el *Random Forest*. Tras escoger este modelo se ha llevado a cabo una exploración de las variables más relevantes y con ella se ha descubierto una mejora en el modelo. El uso de la función de normalización en este modelo provoca que se pierda la relación entre el precio y una variable muy importante como es la demanda.

Por último, se realiza una propuesta de diseño para la construcción de un producto que sirva para la puesta en producción así como una propuesta de explotación de un producto como este para su posible comercialización en el mundo real. En este diseño se ha tenido en cuenta las tecnologías más punteras, así como aspectos de eficiencia en el desarrollo y el despliegue.

1.3. Estructura del documento

La memoria de este trabajo se ha dividido en los siguientes apartados:

- En el Capítulo 2 se trata el funcionamiento del mercado eléctrico con especial incidencia

en cómo se calculan los precios en el mercado. Además este capítulo incluye un detallado análisis de la serie temporal del precio.

- En la siguiente sección, el Capítulo 3, se explican los fundamentos teóricos en los que se asienta el trabajo realizado así como las técnicas usadas en el pasado por diferentes autores con el mismo fin que tiene esta investigación, la predicción de precios del mercado eléctrico.
- En el Capítulo 4 se describe el procedimiento seguido para la creación de los distintos modelos y su entrenamiento, las métricas de error tenidas en cuenta, se realiza un análisis de los errores obtenidos por el mejor modelo y se realiza una propuesta de explotación del modelo predictivo.
- Por último, también existen apartados de referencias, un apartado de acrónimos y otro para anexos, donde se incluyen las tablas con los resultados de los modelos con más detalle.

2

Sistema eléctrico

“I’ve found out so much about electricity that I’ve reached the point where I understand nothing and can explain nothing.”

Pieter van Musschenbroek

El mayor reto a la hora de desarrollar un proyecto de Ciencia de Datos suele ser comprender el marco que rodea al problema al que nos enfrentamos. Este caso no lo es menos; de hecho, pese a que la energía eléctrica desempeña un papel fundamental en la sociedad actual y no se podría imaginar una sociedad sin el uso de esta energía, el funcionamiento del sector es bastante complicado ya que hay muchos actores involucrados con intereses enfrentados en numerosas ocasiones. Es por esto que para conseguir construir un modelo predictivo eficaz para el precio de la energía necesitamos en primer lugar describir y comprender el funcionamiento del sistema eléctrico.

En este capítulo se realizará una introducción al funcionamiento y organización del sector. En particular se realizará una descripción detallada de todos los participantes e interesados del mercado eléctrico y en particular del cálculo del precio, que es el elemento central en este estudio. De igual forma, para comprender la envergadura del problema y qué necesidades de preprocesado de datos puede tener, se realizará un análisis descriptivo de la serie temporal de precios.

2.1. Organización y características del sistema eléctrico español

El sistema eléctrico español ha sufrido un proceso de transformación mediante una liberalización del mismo que comenzó en 1997. Hasta aquel momento, la actividad del sector estaba concentrada en unas pocas empresas que ejercían un monopolio en las distintas regiones españolas. Desde 1997 hasta nuestros días, todas las labores relacionadas con el suministro de energía eléctrica a toda la sociedad como son la generación, el transporte, la distribución, la comercialización y la gestión económica y técnica del sistema eléctrico, han quedado separadas. Esta división implica que incluso grupos empresariales que contienen compañías en diferentes

apartados del sistema no puedan comunicarse entre sí. Producto de esta fragmentación han surgido empresas especializadas en alguno de los campos que son fundamentalmente:

- **Generación.** Consiste en la producción de energía eléctrica, ya sea a través de medios renovables como parques eólicos, plantas fotovoltaicas, etc. o de medios no renovables como centrales nucleares o combustibles fósiles. Para vender esta energía, las empresas generadoras tienen que ir a los mercados eléctricos actuando ellos mismos como vendedores o bien a través de representantes.
- **Transporte.** Las empresas que se dedican a este campo, tienen como objetivo la transmisión de la energía eléctrica vertida por parte de los generadores a la red de transporte y que será utilizada con el fin de suministrarla a los distintos distribuidores y/o para la realización de intercambios internacionales. En el caso español, la Ley 24/2013 de 26 de diciembre del Sector Eléctrico, atribuye a Red Eléctrica de España (en adelante REE) en régimen de exclusividad el ejercicio de las actividades de transporte y operación del sistema, así como de la función de gestor de la red de transporte. La red de transporte, es la red compuesta por las redes de **alta tensión**, que transportan la electricidad desde las plantas generadoras hasta las subestaciones de transformadores con una tensión que supera los 25 kilovoltios, y **media tensión** que transportan la electricidad desde las subestaciones anteriores hasta las de baja tensión a partir de las cuales se suministra la corriente a todos los hogares y que tienen una tensión comprendida entre 1 y 25 kilovoltios.

Por tanto, como la gestión técnica del sistema eléctrico está encomendada al operador del sistema, en España se encarga REE, a quién corresponde la realización de todas aquellas funciones que se derivan del funcionamiento de los servicios de ajuste del sistema, de las desviaciones producidas en el mercado de producción de energía eléctrica, así como la liquidación y comunicación de las obligaciones de pago y derechos de cobro a que den lugar los servicios de ajuste del sistema y la garantía de potencia.

- **Distribución.** La actividad de distribución consiste en la transmisión de la energía eléctrica desde las redes de transporte, o en su caso desde otras redes de distribución o incluso desde la generación conectada directamente a la propia red de distribución, hasta los puntos de consumo u otras redes de distribución con el fin de suministrarla a los consumidores finales (hogares y empresas). Algunos de los principales distribuidores en **baja tensión**, menos de 1 kilovoltio, son: Endesa, Iberdrola, Gas Natural Fenosa (GNF), Energía De Portugal (EDP) y Viesgo.
- **Comercialización.** Las empresas comercializadoras de energía eléctrica acceden a las redes de transporte o distribución, pagan una tarifa por ello y tienen como función la venta de energía eléctrica a los consumidores a través de contratos de suministro. Para ello compran la energía en el mercado eléctrico mayorista, que como veremos más adelante está regulado por el OMIE (Operador del Mercado Ibérico de Energía - Polo Español) y, como ya se ha referido anteriormente, deben pagar unas cuotas a las empresas distribuidoras por el uso de su red. Al contrario que en el transporte y la distribución, en la comercialización existen multitud de empresas, la mayor parte de ellas creadas en los últimos años, encargadas de realizar esta labor en España.
- **Gestión económica y mercado eléctrico.** El mercado eléctrico español es el conjunto de mercados donde se negocia la compra y venta de la mayor parte de la energía eléctrica que se consume en el estado español. Igual que la gestión técnica ha sido encargada a REE, la gestión del mercado eléctrico en España se encuentra dividido en los mercados a plazo,

donde se fijan precios para compra-venta de energía a futuro, y los mercados *spot*, diario e intradiario, donde se negocia la energía para el día siguiente. Esta división de tareas se acordó en el Convenio internacional relativo a la constitución de un mercado ibérico de la energía eléctrica entre España y Portugal firmado el 1 de octubre de 2004 en Santiago de Compostela. Las dos empresas que regulan estas tareas son:

- El OMIE, es responsable de realizar la casación de las ofertas de compra y venta en los mercados diario e intradiarios, así como de la liquidación y comunicación de los pagos y cobros a que dé lugar la energía contratada en los citados mercados dependiendo del precio de casación.
- Por su parte, el OMIP (Operador del Mercado Ibérico de Energía - Polo Portugués, S.A.) controla el mercado a plazo.

Además de los mercados diario e intradiarios y las contrataciones a plazo, existen otro tipo de operaciones que conllevan transacciones económicas como las contrataciones bilaterales, acuerdos particulares entre empresas para la compra-venta de energía, y la aplicación de penalizaciones y recompensas por los servicios de ajuste del sistema y de las desviaciones que se produzcan en el mercado.

En este trabajo pondremos el foco sobre los mercados diario e intradiario, detallados de forma exhaustiva en las siguientes subsecciones, y no tanto en las contrataciones bilaterales y a plazo ya que la forma de calcular los precios es totalmente distinta para los distintos tipos de transacciones y son los precios del mercado diario para los que se quiere obtener un modelo predictivo preciso.

En resumen, se puede decir que en el sistema eléctrico español, todo el proceso que ocurre en el sector energético para ofrecer el suministro eléctrico es bastante complejo quedando separado por un lado el flujo eléctrico y por otro el flujo económico. Todo este proceso queda esquematizado en la figura 2.1.

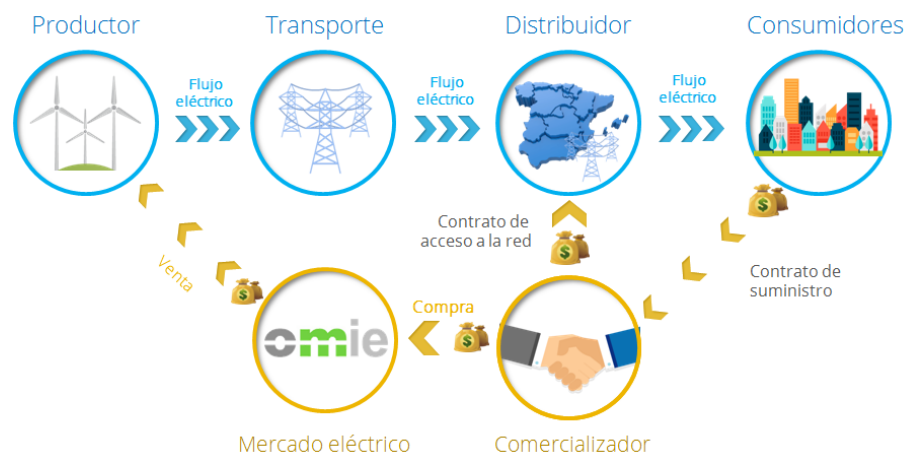


Figura 2.1: Funcionamiento del sistema eléctrico español. *Elaboración propia.*

2.1.1. Mercados *spot*

El mercado *spot* de energía eléctrica proporciona a los agentes participantes la posibilidad de contratar energía eléctrica en siete sesiones: la primera y principal, el mercado diario, y seis

sesiones posteriores, pertenecientes al llamado mercado intradiario, distribuidas a lo largo de las 24 horas del día. Como se puede ver en la figura 2.2, las horas de apertura de las sesiones del mercado intradiario son 17:00, 21:00, 01:00, 04:00, 08:00 y 12:00, para las sesiones de la primera a la sexta respectivamente. La duración del periodo de recepción de ofertas es en general de cuarenta y cinco minutos, excepto en el primer intradiario que es de ciento cinco minutos, comenzando a las 17:00 y finalizando a las 18:45. Por otro lado, el mercado diario se cierra a las 12:00 y Red Eléctrica de España está obligada a publicar información útil como la previsión de demanda, la situación de la red de transporte, las indisponibilidades, las capacidades de importación y/o exportación a las interconexiones internacionales y las ejecuciones de contratos bilaterales antes de las 10:30.

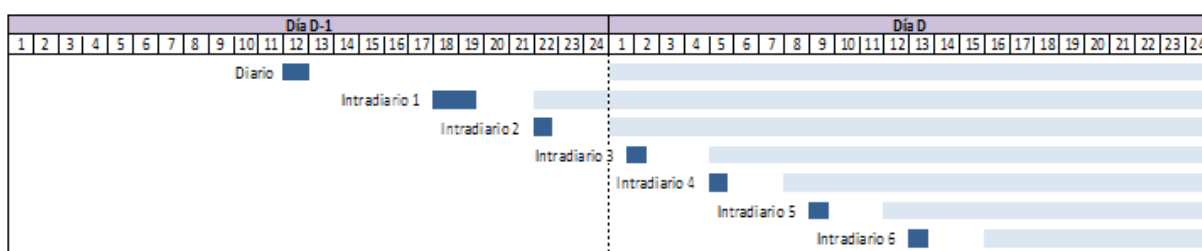


Figura 2.2: Horarios de los mercados de energía eléctrica. *Elaboración propia.*

Este formato de mercados es probable que quede obsoleto en un futuro cercano pues se pretenden eliminar los mercados intradiarios por un mercado continuo en el que se negocia la energía de forma horaria.

Agentes del mercado

Los sujetos del mercado eléctrico son las empresas habilitadas para actuar directamente en él como vendedores y/o compradores de electricidad. Pueden participar como sujetos del mercado los productores, distribuidores y comercializadores de electricidad así como los consumidores de energía eléctrica y representantes de cualquiera de los anteriores.

La denominación de agentes del mercado diario de producción se reserva para los sujetos del mercado que participen en los mercados diario o intradiario. Por tanto, los sujetos del mercado pueden acudir al mismo como agentes del mercado diario o celebrar contratos bilaterales. Estos contratos bilaterales son contratos privados, normalmente entre una empresa generadora y otra comercializadora o un gran cliente final por la cual se comprometen a vender/comprar una cierta cantidad de energía a un precio determinado. Una vez declarados estos contratos bilaterales, su ejecución pasa a ser firme con los mismos derechos y obligaciones que las transacciones del mercado organizado.

Mercado diario

El mercado diario es el más importante porque es donde se negocia la mayor parte de la energía. En él participan todas las unidades de producción en régimen ordinario compuestas por tecnologías como el carbón, centrales nucleares y grandes centrales hidroeléctricas, las unidades de producción en régimen especial, donde se encuentran el resto de tecnologías que lo deseen. Estas unidades de producción podrán participar por la parte de potencia no vinculada previamente

a otra modalidad de contratación (bilateral) puesto que en tal caso se estaría vendiendo la energía dos veces. Por otro lado, también participan de este mercado los comercializadores, los agentes externos autorizados, los distribuidores por la energía neta necesaria, y los grandes consumidores también denominados consumidores cualificados. Los agentes realizan las ofertas de compra/venta de la siguiente manera según sus necesidades:

- Las unidades de producción acceden al mercado para ofertar la venta de sus excedentes de electricidad disponibles que no estén vinculados a un contrato bilateral o para la adquisición de la energía que precisen.
- Las ofertas de compra las presentan los titulares de unidades de adquisición, bien sean comercializadores, distribuidores, consumidores, agentes externos o titulares de centrales de bombeo.
- Los agentes externos y los comercializadores que intuitivamente solo deberían estar interesados en la compra, también pueden presentar ofertas de venta en caso de que estén representado a terceros o que tengan excesos de compra de energía por otras vías, como los contratos bilaterales.

Los distintos agentes, ya mencionados, realizan ofertas de compra y de venta en el mercado que al menos contendrán la cantidad ofertada y el precio al que la quieren comprar o vender. Las ofertas simples se presentan para cada periodo horario y unidad de producción teniendo que contener al menos un precio y una cantidad de energía a comprar (o vender).

Las ofertas complejas son aquellas que, cumpliendo con los requisitos exigidos para las ofertas simples, incorporan, además, todas o algunas de las siguientes condiciones técnicas o económicas:

- La **condición de indivisibilidad** permite fijar un valor mínimo de funcionamiento. Por tanto, esta condición establece que si un tramo indivisible de la oferta resulta casado, lo sea por toda la energía ofertada y no por una fracción de la misma, salvo en lo establecido en la aplicación de las reglas de reparto.

Los vendedores sólo pueden incorporar a la oferta de venta de energía por cada unidad de venta la condición de indivisibilidad para el primer tramo de oferta de los 25 tramos de capacidad de producción posibles en cada periodo de programación.

Los vendedores solo pueden incorporar la condición de indivisibilidad en aquellas ofertas que no declaren ninguna otra condición compleja.

- La **condición de ingresos mínimos** permite a los vendedores incluir como condición en las ofertas de venta de energía que presenten que dicha oferta sólo se entiende presentada a los efectos de la casación si obtiene unos ingresos mínimos para el conjunto de su programación en el día. Estos se expresarán como una cantidad fija, declarada en euros, y como una cantidad variable declarada en euros por MWh.

La condición de ingresos mínimos no podrá ser tal que el ingreso solicitado supere en más de un 100 % al ingreso resultante de la aceptación completa de la oferta al precio ofertado.

- La **condición de parada programada** permite que si la unidad de producción ha sido retirada de la casación por no cumplir la con la condición de ingresos mínimos, esta realice una parada en un máximo de tres horas mediante la aceptación de ese tramo de tres horas como ofertas simples. La energía ofertada que incorpore la condición de parada programada

deberá ser decreciente durante los periodos de programación para los que se declara la condición.

En todo caso, las ofertas rechazadas por la condición de ingresos mínimos que tienen la condición de parada programada son también no divisibles, salvo en lo establecido en las reglas de reparto, sin que pueda existir ningún otro tramo de producción no divisible en el mismo periodo de programación.

- **Condición de gradiente de carga:** consiste en establecer para cada unidad de venta una diferencia máxima de variación de energía entre dos periodos de programación consecutivos para evitar cambios bruscos en las unidades de producción que en algunos casos con tecnologías antiguas no pueden ser asumidos.

Una vez celebrada la sesión del mercado diario, el operador del sistema estudia la viabilidad técnica del programa de funcionamiento para garantizar la seguridad y fiabilidad del suministro. Si el programa resultante del mercado diario más los contratos bilaterales no cumplen los requisitos de seguridad, el operador del sistema resuelve dichas restricciones técnicas modificando el programa de las unidades de producción resultante del mercado diario y de los contratos bilaterales.

Mercado intradiario

El mercado intradiario consta actualmente, como ya hemos comentado, de seis sesiones que se celebran a lo largo de las 24 horas del día, pudiendo acudir al mismo, como compradores y vendedores, todos los agentes del mercado diario que hubieran participado en la sesión correspondiente de dicho mercado diario. También, excepcionalmente, pueden presentar ofertas las unidades de producción que hubieran comunicado la indisponibilidad al operador del sistema con anterioridad al cierre del mercado diario de producción y que hayan recuperado su disponibilidad, así como todos los agentes, aun sin haber participado en el mercado diario pero que previamente hubiesen comunicado al operador del sistema la existencia de un contrato bilateral para las horas comprendidas en la sesión correspondiente del mercado intradiario.

2.2. Cálculo del precio

El precio de la electricidad en el punto de suministro del consumidor tiene varias componentes: el precio del mercado que retribuye la entrega de energía eléctrica, junto con los servicios de ajuste y la tarifa o peaje a satisfacer por el uso de las redes de transporte y distribución, costes por pérdidas, alquiler del equipo de medida e impuestos entre otros. Este trabajo se ha centrado en el primero de todos ellos.

En el año 2009 surgió la iniciativa llamada «*Price Coupling of Regions*» (*PCR*) por parte de siete mercados europeos de electricidad, entre los que se encuentra el español, OMIE. En esta iniciativa se desarrolló un algoritmo llamado *Euphemia*. Este algoritmo calcula los precios de la energía eléctrica de manera eficiente, persiguiendo la maximización del *welfare*, que se define como el excedente o ganancia, tanto de los compradores como de los vendedores, al tiempo que optimiza el uso de la capacidad disponible en las interconexiones.

Para esa maximización del *welfare*, tanto para el mercado diario como para el intradiario, el algoritmo *Euphemia* considera curvas agregadas en escalón, que se corresponden con las curvas

para las que el precio de inicio de aceptación de un tramo de energía y el precio de aceptación completa de dicho tramo de energía es coincidente. De forma resumida, las empresas encargadas de la generación realizan sus ofertas (cantidad de energía y precio) y las empresas encargadas de la comercialización, consumidores directos, etc. demandan la energía necesaria a un precio determinado.

Una vez realizadas las ofertas estas se ordenan en escalón de forma creciente según el precio, en el caso de la venta y en orden decreciente en el caso de la compra y el punto de corte de las curvas de oferta y demanda se denomina punto de casación. En principio ese es el punto que optimiza el *welfare* y, por tanto, establece el precio de la energía para esa determinada hora.

Toda la energía ofertada y demandada a un precio menor al de casación será intercambiada a ese precio, mientras que la que tenga un precio mayor no. Este proceso se repite para cada una de las 24 horas del día.

Sin embargo, la determinación del precio no es tan trivial como la agregación de las curvas de oferta y demanda. Si nos fijamos en un ejemplo concreto del mercado, por ejemplo, en la figura 2.3 se pueden apreciar las curvas de compra/venta agregadas para el día 22 de agosto de 2016. En ellas, podemos observar que el punto de corte entre las curvas agregadas de compra y de venta no son exactamente el precio de la energía para esa hora. Esto ocurre porque algunas ofertas contienen condiciones complejas, además de que existe la posibilidad de importación y exportación de energía con los países fronterizos con España (Marruecos, Portugal y Francia).

Curvas de oferta de compra y venta

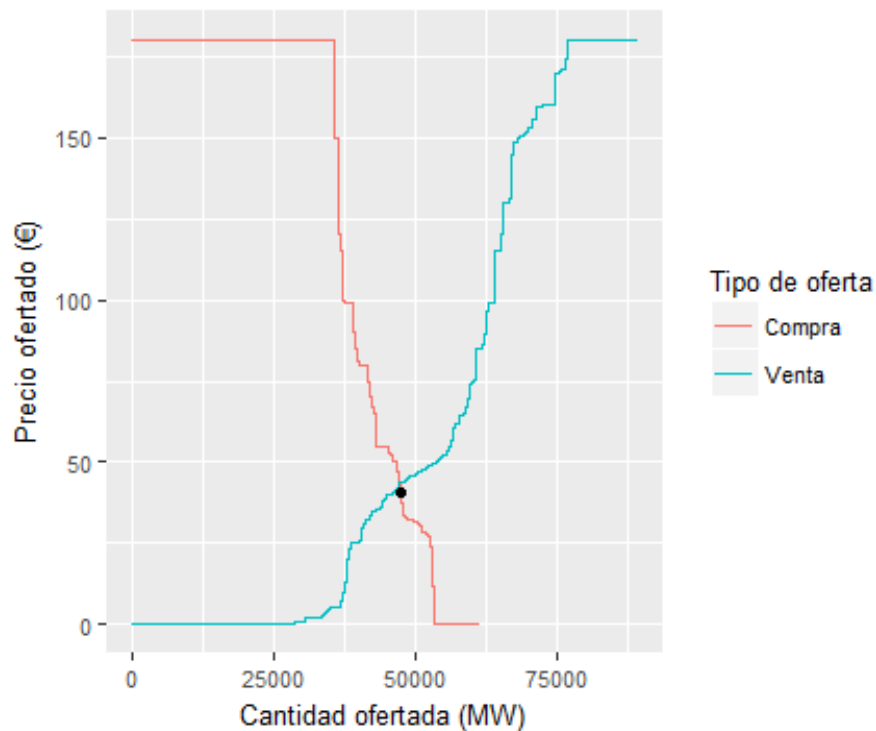


Figura 2.3: Curvas agregadas de compra/venta y precio de la energía para el 22/08/2016 a las 00:00 *Elaboración propia*.

Por tanto, para determinar el precio de casación se introducen las importaciones y exportaciones, así como las condiciones complejas mencionadas anteriormente (de mínimos

ingresos, de indivisibilidad, de parada programada, de variación de capacidad de producción o gradiente de carga). Se comprobará también que cumple las siguientes restricciones:

- El precio final de casación es coherente con las ofertas de compra y venta que se han producido para el periodo en cuestión en ese mercado en particular.
- El precio final de casación está dentro de los límites (si los hubiese) del mercado en cuestión. En el caso español, dichos límites son 0 y 180,3 €.

Toda esta casuística de condiciones complejas e importaciones y exportaciones pueden variar el precio de la energía de manera significativa. Como ejemplos:

- **Condiciones complejas:** en el caso de que no se cumpla con alguna de las condiciones complejas de la oferta, esa oferta será eliminada de la curva agregada de venta y por tanto encarecerá el precio de la energía como podemos ver en la figura 2.4.

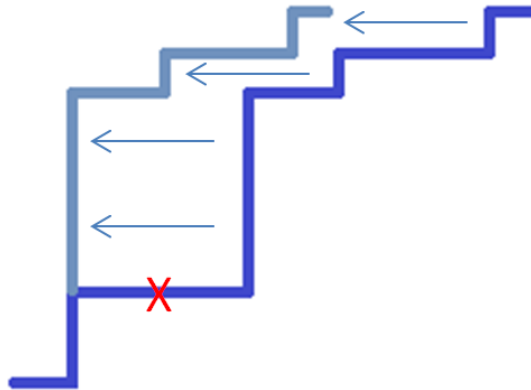


Figura 2.4: Oferta con condiciones complejas no cumplida. *Elaboración propia.*

- **Exportación:** la exportación de energía a otros sistemas eléctricos que se encuentren conectados produce un abaratamiento del precio en el sistema vecino y un encarecimiento en el nacional. Esto es debido a que la curva de compra se modifica ya que hay parte de la oferta que va a parar al país importador. Este efecto es el que podemos observar en la figura 2.5.
- **Importación:** en este caso si importamos energía, la energía nacional se abarata porque existe más oferta de venta desplazando esta curva hacia la derecha como podemos ver en la figura 2.6.

Si nos fijamos en el día 22 de agosto podemos ver que efectivamente hubo importaciones y exportaciones. En concreto aquel día España tuvo un balance importador neto de 1608,4 MWh (se importó 2,250 MWh de Francia y se exportó 199,6 MWh, 42 MWh y 400 MWh a Portugal, Andorra y Marruecos, respectivamente, como se puede ver en la figura 2.7). Estas exportaciones están destinadas a igualar los precios de la electricidad en todos los países aunque estos intercambios están limitados por la capacidad de intercambio para cada hora, que es publicada por REE.

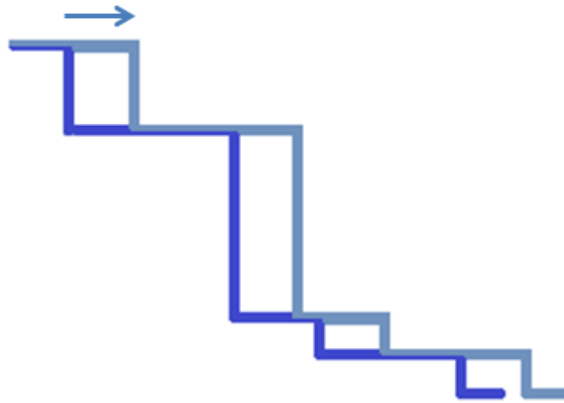


Figura 2.5: Efecto de la exportación en la curva de oferta de compra. *Elaboración propia.*

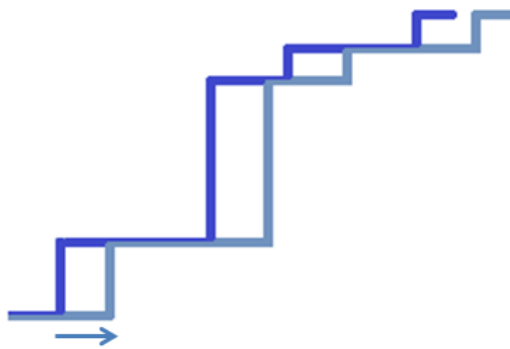


Figura 2.6: Efecto de la importación en la curva de ofertas de venta. *Elaboración propia.*

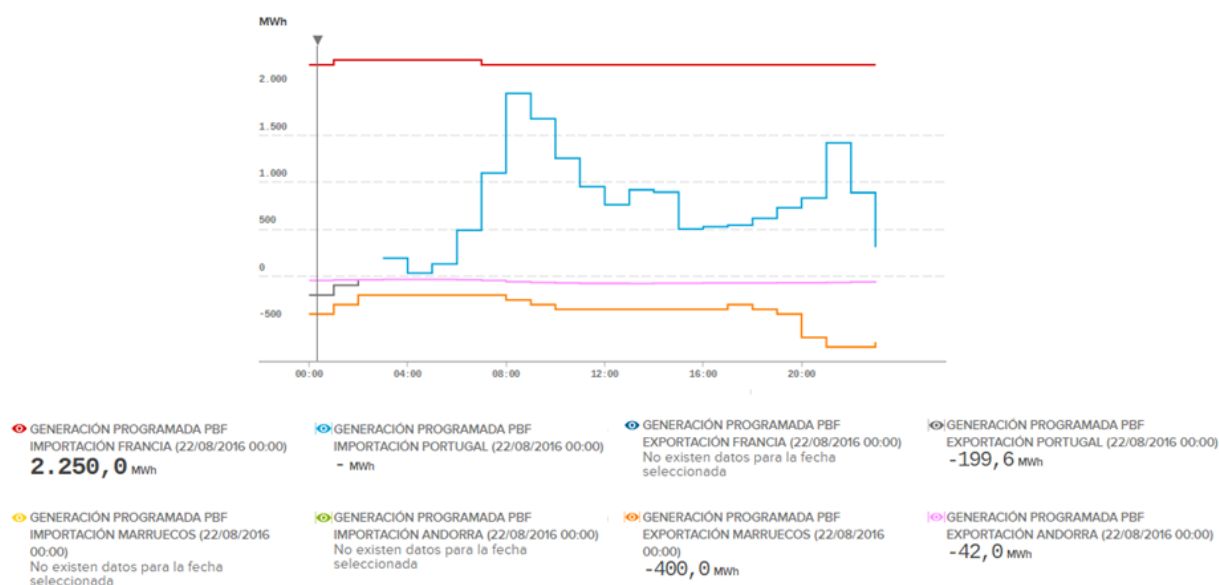


Figura 2.7: Intercambio de energía con los países limítrofes a España. *Fuente REE.*

Si añadimos esas importaciones y exportaciones a la curva de la figura 2.3, de la manera que se ha explicado antes, se produce un desplazamiento en la curva de venta generando la curva de venta casada (verde) de la figura 2.8. De esta manera ahora sí que ocurre que el punto de corte entre la compra y la venta coincide con el precio de casación.

A lo largo de la historia reciente se han producido numerosas situaciones importantes, como subidas o bajadas bruscas del precio de materias primas como el carbón o el gas que son determinantes a la hora de la casación del precio. Precios altos de combustibles, empleados en las centrales térmicas, generarán ofertas de dichas centrales más altas y, por tanto, un aumento en el precio del MWh si es necesario su uso para satisfacer la demanda del país. Recientemente, al comienzo de 2017, por ejemplo, se produjo una situación en la que el precio de la electricidad experimentó un aumento inesperado, copando todos los titulares de prensa: «El precio de la luz se dispara este miércoles en plena ola de frío» [4], «Las cementeras advierten de posibles paros de fábricas por el precio de la luz» [7].

2.3. Análisis de la serie temporal de precios

Los precios de la electricidad en el mercado eléctrico son muy volátiles ya que hay muchos factores que provocan que exista una gran variabilidad en los datos. Como ya hemos visto, existen dos factores clave en la determinación de los precios, que son la demanda y la generación.

En primer lugar, la demanda tiene una variabilidad que se explican en parte mediante factores climáticos, de laboralidad y de rutinas de la población. Por ejemplo, la cantidad de energía consumida aumenta en las horas de trabajo y de llegada a la vivienda, alrededor de las 19:00-20:00, cuando se conectan numerosos aparatos eléctricos, luces, etc. Junto a estos factores existen otros más difíciles de conocer a priori su impacto: eventos, huelgas, etc.

Curvas de oferta de compra y venta

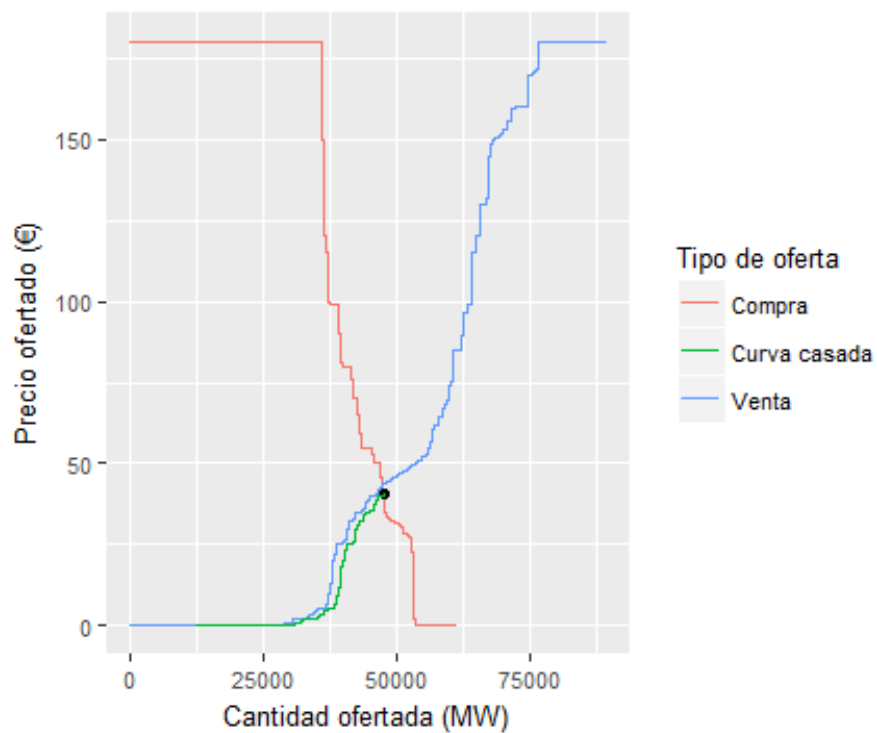


Figura 2.8: Curvas agregadas de compra/venta y precio de la energía tras exportaciones e importaciones y tras condiciones complejas para el 22/08/2016. *Elaboración propia.*

En segundo lugar, en el caso de la generación existe mucha incertidumbre. Las energías renovables en su mayor medida no pueden almacenar la energía producida y existe una variabilidad altísima en función de la meteorología y las condiciones de las plantas generadoras. Además, algunas de las plantas productoras hacen uso del petróleo, el gas o el carbón y, por tanto, producirán a un precio u otro en función del precio que tengan, a su vez, dichos recursos en el mercado.

Por último, existen estrategias de compra/venta de energía al margen del mercado eléctrico mediante contratos bilaterales entre productores y comercializadores y estrategias especulativas en el mercado que pueden producir modificaciones significativas en los precios del mercado diario. También el precio se ve afectado por la demanda y la producción en otros países, en el caso de España normalmente Francia, por el proceso de importación y exportación de energía que antes hemos comentado.

De manera concisa, la predicción del precio de la energía se puede ver afectada por muchos factores que son muy complejos y en algunos casos sobre los cuales no existen datos o son estrategias comerciales complejas de empresas que por supuesto no revelan al público en general.

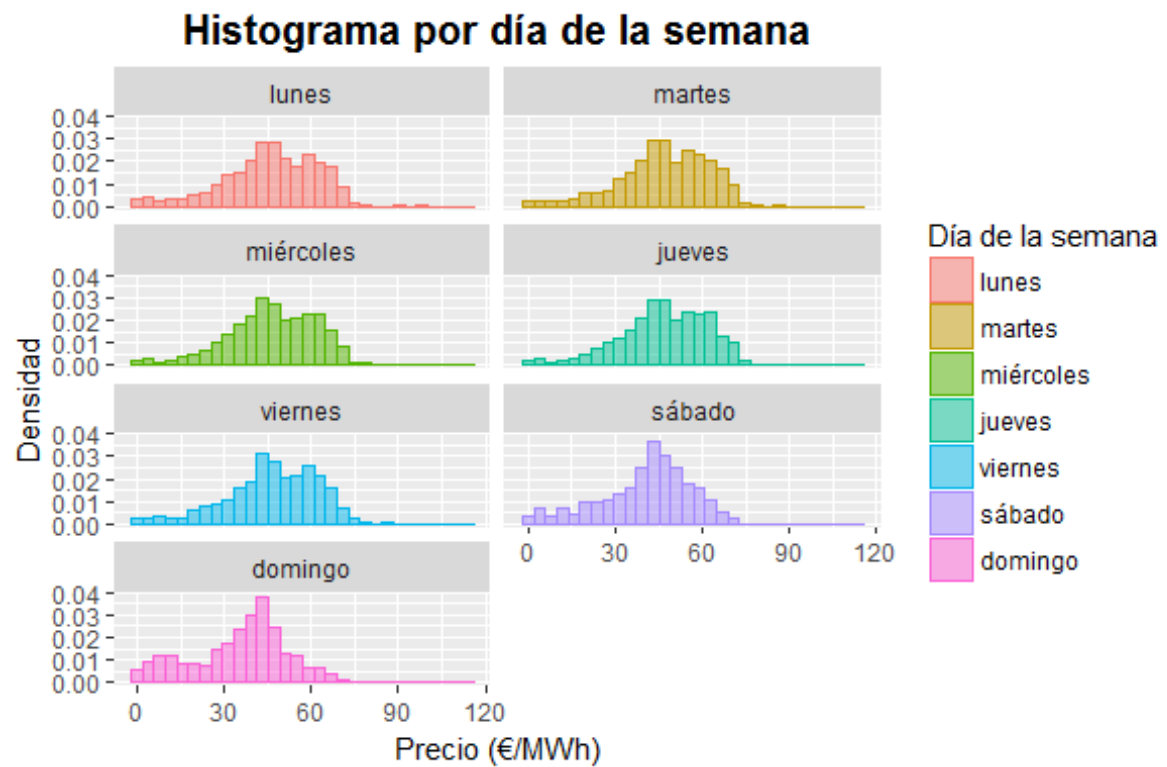
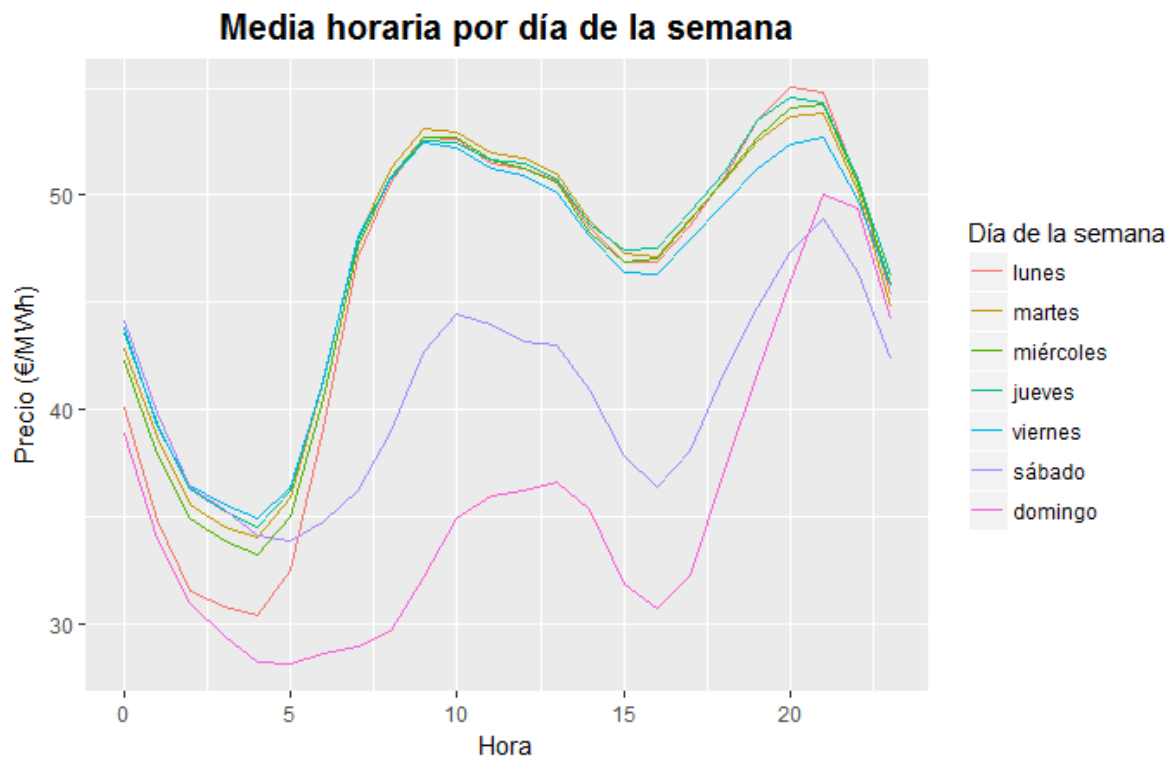
A pesar de las dificultades ya comentadas, los primeros análisis de la serie temporal de precios nos indican una fuerte estacionalidad en los datos debida fundamentalmente al efecto de la demanda ya comentado. Si analizamos el precio en los días de la semana podemos observar claramente en el histograma de la figura 2.9 que los días con una distribución distinta de precios son los días de fin de semana, en los cuales los precios son habitualmente menores. Todos los días poseen la mayor parte de los precios en torno a los 45€ pero en el caso de los días laborables existen un buen número de horas con precios superiores a esos 45€, mientras que en el caso de sábados y domingos hay un número mucho más reducido de esas horas. Además se puede ver un mayor peso de horas con precio por debajo de los 30€ tanto en sábados como en domingos, aunque algo superior en estos últimos.

Si analizamos los precios de forma horaria podemos confirmar el análisis anterior que nos indicaba que había horas más baratas en fin de semana que entre semana. Como ocurría en el caso diario, en el análisis horario los precios siguen la tendencia de la demanda eléctrica. De manera que, durante la noche, existe una menor demanda y, por tanto, los precios son más bajos.

En la figura 2.10, que se ha realizado mediante la media horaria de los precios entre 2014 y 2016, se puede observar que la bajada de los precios comienza a partir de las 21:00 hasta aproximadamente las 5:00 donde los precios comienzan a subir hasta las 9:00, de forma muy pronunciada en los días laborables y hasta las 10:00 y, de forma no tan pronunciada, los fines de semana. Los precios se mantienen hasta las horas de la comida donde empiezan a disminuir. Este valle comprende aproximadamente de 13:00 a 17:00. Por último, los precios alcanzan su mayor cota alrededor de las 20:00-21:00 entre semana y 21:00-22:00 los fines de semana.

Si analizamos los precios de forma mensual como en la figura 2.11, vemos que en los meses de verano los precios reducen su volatilidad horaria y que se mantienen las diferencias entre fines de semana y entre semana a lo largo de los distintos meses. El efecto de la reducción de la volatilidad horaria en los meses de verano es probablemente producida por el efecto de los sistemas de refrigeración, el efecto del turismo y la disminución productiva en el sector industrial.

Por último, en este estudio de precios es interesante observar qué ocurre en los festivos ya que los patrones de consumo eléctrico en estos días de la población en general son claramente distintos a los días laborables. Para llevar a término este análisis, solamente han sido tenidos en cuenta los festivos nacionales y de Comunidad Autónoma, que se encuentran publicados en el

Figura 2.9: Histograma de precios diferenciado por día de la semana. *Elaboración propia.*Figura 2.10: Media horaria por día de la semana. *Elaboración propia.*

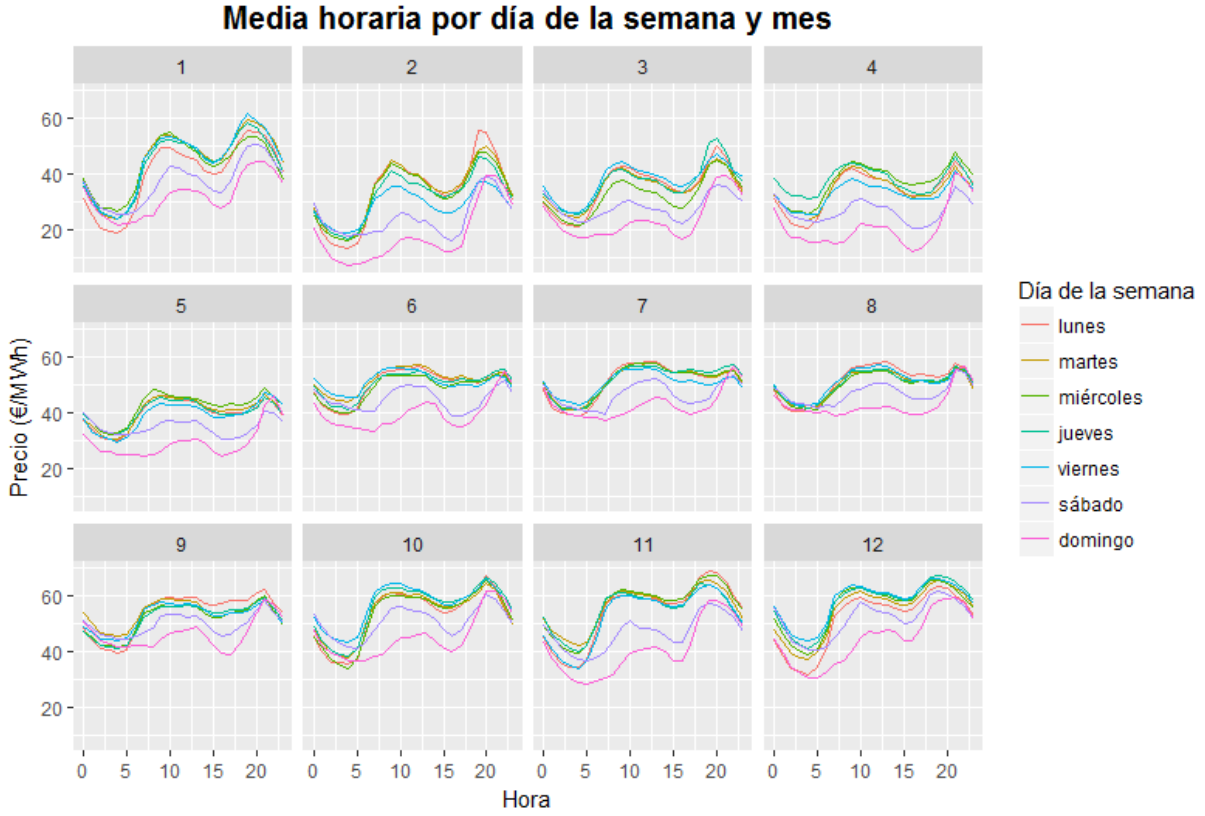


Figura 2.11: Media horaria por día de la semana y mes. *Elaboración propia.*

BOE, pero no así aquellos relativos a los ayuntamientos puesto que las fiestas particulares de un ayuntamiento afectan al consumo de ese ayuntamiento pero ese efecto desaparece al agregarlo con el resto de la población española. Como uno de los productos de este análisis se ha definido el coeficiente de festividad para el día i como el porcentaje de población en España en fiesta el día i en tanto por uno. Es decir,

$$CF_i = \frac{\sum_{C \in \Omega_{fest_i}} P_C}{\sum_{C \in \Omega} P_C} \quad (2.1)$$

donde Ω es el conjunto de Comunidades Autónomas y P_C la población de la Comunidad Autónoma C .

Este coeficiente nos permite ver, en la figura 2.12, las diferencias entre los días festivos ($CF \neq 0$) y los días laborables ($CF = 0$ y día de la semana distinto de domingo) y los domingos (en negro). Se observa que los festivos tienen unos precios menores a los días laborables y que en general cuanto mayor es el coeficiente de festividad y, por tanto, mayor porcentaje de la población española se encuentra de vacaciones, menor es el precio.

El análisis de la variable objetivo llevado a cabo es muy importante, tanto para la comprensión del problema como para tomar ciertas decisiones en la implementación que se detallarán en el Capítulo 4 sobre todo en la fase de preprocesado de los datos. En concreto, en ese capítulo se realiza una normalización de los precios siguiendo los resultados del análisis de la relación

entre laboralidad y precio y para reducir la varianza de los datos y que es fundamental para los resultados finales.

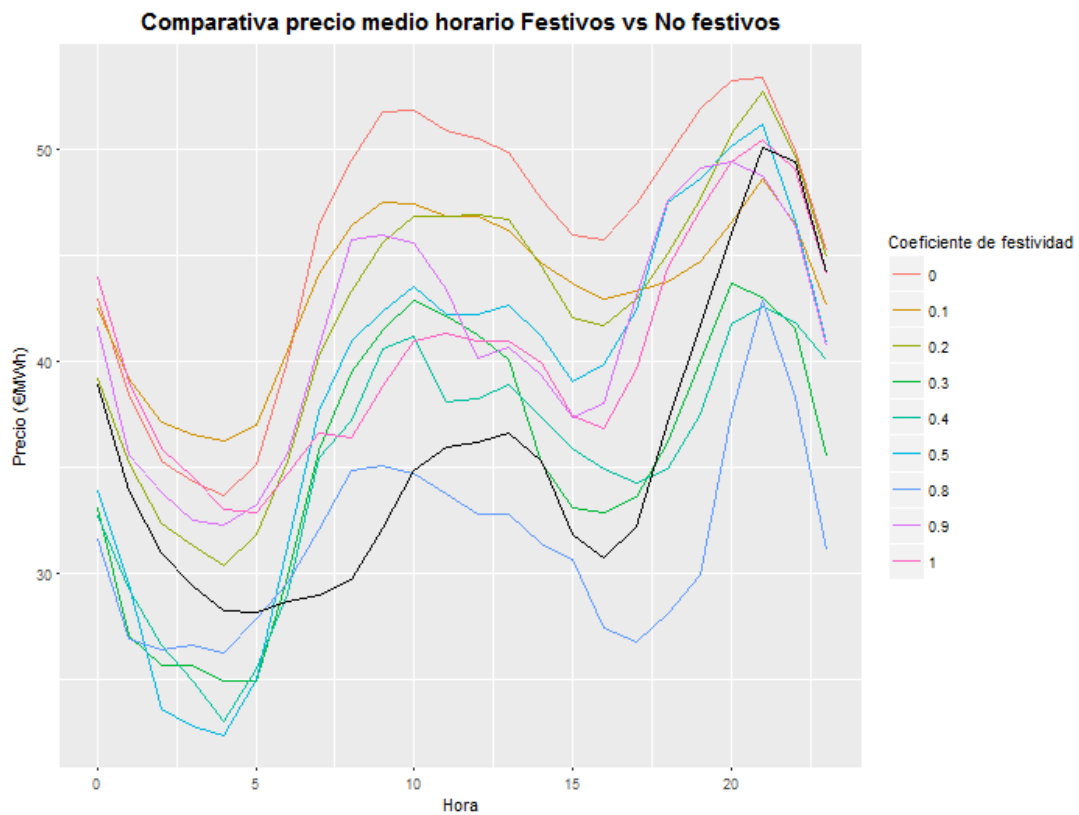


Figura 2.12: Comparativa del precio medio horario en festivos contra no festivos. *Elaboración propia.*

3

Estado del Arte

“Mathematics is the music of reason.”

James Joseph Sylvester

Este capítulo cubre el marco teórico de algunas de las técnicas más usadas en problemas de predicción, más en concreto, aquellas que han sido utilizadas recientemente en el ámbito de la predicción del precio de la energía y el uso de estas técnicas en dicho problema.

Como se pueden observar en [13] y en [48] las aproximaciones tomadas por distintos autores en el pasado para el problema de la predicción del precio de la energía han sido diversas. Estos métodos van desde la teoría de juegos a la inteligencia computacional pasando por modelos de simulación y de series temporales. Es muy descriptivo el esquema de la figura 3.1 usado en [48] para describir las distintas alternativas y que va ser seguido con alguna modificación en esta revisión.

En nuestro estudio nos centraremos sobre todo en la parte derecha del esquema, es decir, las técnicas relacionadas con las series temporales y la inteligencia computacional. Esto es debido a que estas técnicas son las que se encuentran mucho más relacionadas con la propuesta que se lleva a cabo en este Trabajo de Fin de Máster, que a su vez ha sido apoyada por investigaciones como [19] donde se indican que los métodos económicos y de teoría de juegos son una buena aproximación pero ciertamente insuficiente en caso de querer realizar predicciones diarias precisas.

3.1. Modelos de Series Temporales

Las series temporales son sucesiones de valores espaciados en el tiempo de forma constante, es decir, se observa el fenómeno en instantes tomados de manera regular. El análisis

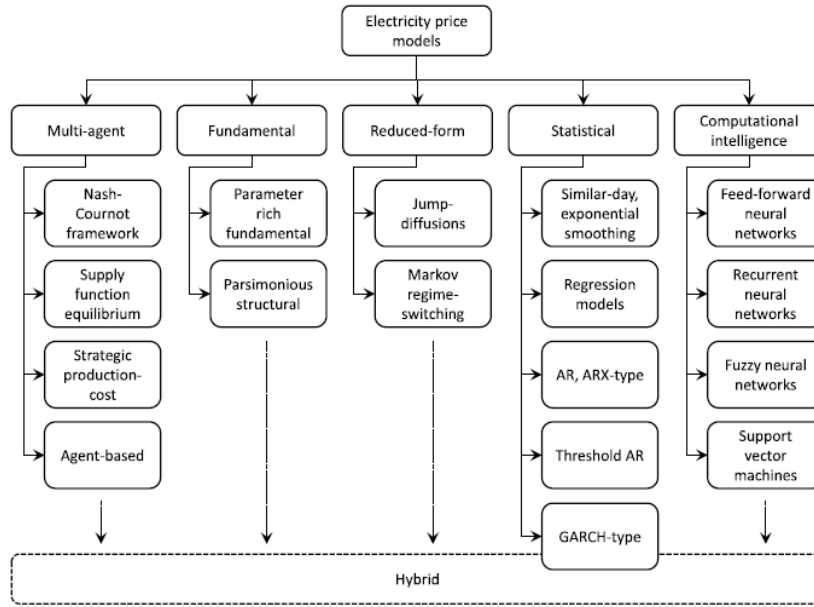


Figura 3.1: Esquema de los modelos de predicción para precios de energía. *Fuente [48]*

de series temporales pretende modelar la estructura temporal subyacente existente en las observaciones tomadas en un determinado período de tiempo. Una vez realizada la modelización, estos algoritmos sirven para comprender la serie temporal pero también para predecir el comportamiento en el futuro de la misma.

Todas estas técnicas han sido ampliamente estudiadas con numerosas publicaciones y libros al respecto como [44] o [25]. Son usualmente aplicadas en los casos en los que la propia serie temporal tenga suficiente información en sí misma como para no necesitar variables exógenas para su comprensión y para su análisis predictivo. El caso de la predicción de precios como ya hemos observado en la Sección 2.3 es uno de esos casos en los que la serie temporal tiene una estructura muy fuerte y, por tanto, el uso de este tipo de análisis es muy adecuado pese a sus limitaciones.

En este apartado recopilaremos algunos de los métodos que habitualmente más se utilizan como aproximaciones para series temporales univariantes como es el precio de la energía. Entre estos métodos se encuentran los modelos autorregresivos (AR), los modelos de media móvil (MA) y los modelos conjuntos como ARMA y ARIMA.

3.1.1. Características de las Series Temporales

Antes de describir los distintos métodos para series temporales es necesario realizar definiciones sobre la nomenclatura a usar y conceptos relacionados.

Una serie temporal puede ser definida matemáticamente como una secuencia de variables aleatorias x_1, x_2, x_3, \dots , donde la variable x_t representa el valor tomado por la variable en el momento t . Por tanto, en general, una serie temporal es un proceso estocástico, es decir colección de variables aleatorias que varían a lo largo del tiempo.

Hay dos conceptos que son fundamentales a la hora de describir una serie temporal: el de

autocorrelación y el de serie estacionaria. Ambos conceptos serán expuestos en las dos siguientes subsecciones.

Series Estacionarias

Un concepto muy importante dentro de las series temporales es la noción de regularidad que puede ser definida mediante dos propiedades de las series temporales: que sean estrictamente estacionarias y que sean débilmente estacionarias.

Una serie temporal es **estrictamente estacionaria** si el comportamiento probabilístico de cada colección de valores

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\} \quad (3.1)$$

es idéntica a la serie temporal desplazada

$$\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}. \quad (3.2)$$

Matemáticamente:

$$P\{x_{t_1} \leq c_1, x_{t_2} \leq c_2, \dots, x_{t_k} \leq c_k\} = P\{x_{t_1+h} \leq c_1, x_{t_2+h} \leq c_2, \dots, x_{t_k+h} \leq c_k\}. \quad (3.3)$$

Como la definición anterior es demasiado fuerte y es muy difícil comprobar que una serie temporal cumpla esa condición se suele imponer la siguiente condición que es mucho menos estricta. Una serie temporal x_t es **débilmente estacionaria** si es un proceso de varianza finita que cumple:

- La media μ_t es constante y no depende del tiempo t
- La autocovarianza $\gamma(s, t)$ depende únicamente de s y t solo a través de la diferencia $|s - t|$

Si una serie temporal cumple esta definición, habitualmente se denomina simplemente **serie estacionaria**.

Autocovarianza y autocorrelación

En muchas ocasiones no conocemos la dependencia entre los valores dentro de una serie temporal pero esta dependencia puede ser medida mediante la covarianza y la correlación como en la estadística clásica pero usando dos momentos dados x_s y x_t de la propia serie temporal puesto que, como ya se ha descrito anteriormente, x_s y x_t son dos variables aleatorias.

Por tanto, podemos definir la función de autocovarianza como el producto del segundo momento

$$\gamma_x(s, t) = cov(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)] \quad (3.4)$$

para todo s y t . De igual forma que con la covarianza de la estadística clásica,

$$\gamma(t, t) = var(x_t). \quad (3.5)$$

Como en la estadística clásica, normalmente es más conveniente el uso de una medida que se encuentre entre -1 y 1 y, por tanto, obtenemos la autocorrelación. La función de autocorrelación

(ACF) se define como:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}. \quad (3.6)$$

De igual modo que en la estadística clásica, este coeficiente nos aporta una medida de predictibilidad lineal de la serie en un tiempo dado t , haciendo uso de otro s .

3.1.2. Modelos autorregresivos

El concepto de **autorregresividad (AR)**, introducido por Yule en 1927 [52], proviene intuitivamente de que la serie temporal estudiada Y_t , es dependiente de valores ocurridos en el pasado ponderados por unos ciertos pesos de la misma variable. Es decir, como veremos en la Subsección 3.2.1, la serie temporal puede ser expresada como una Regresión Multilineal de valores pasados.

Los procesos autorregresivos de orden p , denotados como AR(p), pueden ser expresados matemáticamente de la forma descrita en la ecuación (3.7)

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (3.7)$$

donde ϵ_t es una variable que representa ruido blanco. El ruido blanco es un proceso estocástico que se caracteriza por el hecho de que sus valores en dos tiempos diferentes no guardan correlación estadística; en concreto podemos asumir, aunque no de forma necesaria, que este término es una variable aleatoria tomada de una muestra con distribución normal de media cero y varianza σ_ϵ^2 , es decir, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.

Cabe esperar que estos métodos sean de utilidad para el problema de predicción de precios de la electricidad en España porque como hemos visto en la Sección 2.3, la serie temporal de estos precios tiene una gran componente estacional y hay ciertas horas que siempre son mucho más caras que otras.

3.1.3. Modelo de media móvil

El modelo de **Media móvil (MA)** especifica que la serie temporal Y_t depende linealmente del ruido blanco actual y los q ruidos blancos anteriores. Por tanto, en este caso se calcula como una combinación lineal de estos ruidos blancos. También fue introducido por Yule en 1921 [51].

Matemáticamente, los modelos de media móvil de orden q , denotados por MA(q) cuando la serie temporal Y_t está centrada en 0, pueden ser expresado como se observa en la ecuación (3.8).

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (3.8)$$

donde, de nuevo, ϵ_t es una variable que representa ruido blanco.

3.1.4. Modelos ARMA

En general, no es necesario elegir exclusivamente entre los modelos AR y MA. Estos dos modelos pueden ser combinados para obtener un modelo autorregresivo de media móvil, denotado

por ARMA(p,q). Estos modelos fueron popularizados por Box y Jenkins en 1976 [16]. Este proceso puede ser expresado matemáticamente por la ecuación (3.9).

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}. \quad (3.9)$$

Para hacer uso de todos estos modelos (tanto los AR, como los MA, como los ARMA) de forma adecuada, las series deben ser estacionarias. En muchas ocasiones, las series no lo son pero se pueden usar transformaciones para conseguir convertirlas en series estacionarias. Para ello, usualmente se realiza en primer lugar una eliminación de la tendencia y en segundo lugar una diferenciación de la serie temporal, es decir, computar las diferencias de primer, segundo u orden d . Por ejemplo, en la ecuación (3.10) se muestran las diferencias de primer orden.

$$\Delta Y_t = Y_t - Y_{t-1}. \quad (3.10)$$

Se pueden conseguir modelos más generales al modelo ARMA de manera que se realicen las diferencias comentadas. Estos modelos serán explicados con más detalle en las dos secciones siguientes.

3.1.5. Modelos ARIMA

Con el fin de solventar el problema de los procesos no estacionarios se define un nuevo modelo que realiza las diferencias comentadas en el apartado anterior, **Modelo Autorregresivo Integrado de Media Móvil**, denotado por ARIMA(p,d,q). Matemáticamente puede ser expresado como se encuentra descrito en la siguiente ecuación:

$$Y_t = -(\delta^d Y_t - Y_t) + \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (3.11)$$

donde d corresponde a las d diferencias necesarias para convertir la serie original en estacionaria.

3.1.6. Modelos SARIMA

En ocasiones, la diferencia con un retraso de 1 no es suficiente para convertir la serie en estacionaria. En particular la serie temporal del precio de la electricidad o la demanda eléctrica necesitan diferencias con un mayor retraso. En estos procesos se deben aplicar los denominados **Modelos ARIMA Estacionales (SARIMA)** que contienen una parte no estacional y otra parte estacional. Estos modelos son capaces de modelizar patrones periódicos mensuales, semanales, diarios, etc.

3.1.7. Usos de modelos de series temporales para predicción de precios en la literatura

En la bibliografía existen muchas referencias en las que se usan estos modelos para la predicción de precios como son [26] que usa datos de *Leipzig Power Exchange* para sus experimentos y como modelos AR y ARMA con alguna modificación. En [23] se hace una de las primeras predicciones del precio de la electricidad en España tras el cambio regulatorio y también usa datos del mercado californiano. Hace uso de un modelo *Seasonal ARIMA* con distintos parámetros para los dos distintos mercados. En [22] y en [32] se propone una mejora

usando una transformación *wavelet* previa a la utilización del modelo ARIMA para reducir la volatilidad de la serie temporal que se aplica sobre datos de precios de electricidad; en general, esta transformación ofrece unos mejores resultados que un ARIMA tradicional. También son numerosos los estudios que combinan modelos (S)ARIMA con otros modelos de predicción, que usan variables exógenas [50] o que hacen un modelo por día de la semana o incluso por hora.

3.2. Modelos de aprendizaje automático

El Aprendizaje Automático es un campo de la Inteligencia Artificial cuya definición es compleja y en la cual los diferentes autores no se encuentran de acuerdo pero podríamos definirlo como la materia que estudia las técnicas y algoritmos que permiten a las máquinas adaptarse a situaciones dinámicas y, por tanto, de algún modo aprender a predecir el futuro, a partir de descubrir patrones subyacentes en los datos.

Los modelos de aprendizaje automático se pueden clasificar en algoritmos supervisados, no supervisados y semisupervisados, según si los datos están completamente etiquetados, sin etiquetar o etiquetados parcialmente.

Más detalladamente, el objetivo de los modelos supervisados consiste en aprender, a partir de una matriz de datos X y una variable a predecir y asociada, un modelo que permita realizar predicciones ante nuevos casos, para los cuales ya no conocemos el valor y , de manera lo más precisa posible.

En los algoritmos no supervisados al no tener etiquetas en la matriz de datos, el sistema hace uso de la estructura de los datos para inferir cierto conocimiento de ellos como, por ejemplo, separar los registros en distintos grupos.

Por último, los algoritmos semisupervisados combinan las dos técnicas anteriores dado que tiene algunos datos etiquetados pero que al ser muy pocos con respecto al resto de la muestra, también necesitan usar la estructura subyacente de los datos no etiquetados para inferir conocimiento y aplicarlo a los nuevos datos.

El foco de este trabajo se encuentra en los algoritmos supervisados, más concretamente, el caso de la predicción de precios pertenece a un tipo de modelos supervisados denominados modelos de regresión porque la etiqueta a predecir es un valor real, en este caso que va desde 0 hasta 180,3€ que, como vimos en el Capítulo 2, son el valor mínimo y el valor máximo, respectivamente, que alcanzan los precios en el mercado español.

Algunas de las técnicas más usadas para regresión son la Regresión Lineal Múltiple, los Árboles de Decisión, los K-Vecinos Próximos, las Máquinas de Vectores Soporte, las Redes Neuronales y métodos de combinación o *ensembles* que usan algunos de los modelos anteriores de forma conjunta para conseguir una mejor predicción, entre los que destacan *Random Forest*, *Bagging* y *Adaboost* entre otros.

3.2.1. Regresión Lineal

El modelo de Regresión Lineal Múltiple relaciona una variable regresora e independiente y , con una serie de variables dependientes, explicativas o exógenas, x_i como una combinación lineal

de las mismas más un término de error u , es decir,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u. \quad (3.12)$$

El problema a resolver, es estimar los parámetros fijos y desconocidos:

$$\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_k \quad (3.13)$$

dada una muestra:

$$(y_i, x_{1i}, x_{2i}, \dots, x_{ki}) : i = 1, 2, \dots, n. \quad (3.14)$$

Expresado para todas las observaciones obtenemos el sistema de ecuaciones:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_k x_{k1} + u_1, \\ y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_k x_{k2} + u_2, \\ &\dots \\ y_n &= \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_k x_{kn} + u_n. \end{aligned} \quad (3.15)$$

Matricialmente:

$$\vec{y} = X\vec{\beta} + \vec{u}. \quad (3.16)$$

La matriz X , denominada matriz de datos o regresores, incluye una columna de unos para el término independiente. Dados unos coeficientes β , podemos calcular el error cuadrático medio de los mismos

$$S = \sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (\hat{y}_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \cdots - \hat{\beta}_k x_{ki})^2. \quad (3.17)$$

Si queremos minimizar ese error de predicción se debe resolver el siguiente problema:

$$\min_{\beta} \|X\beta - y\|_2^2. \quad (3.18)$$

Una solución explícita, denotada $\hat{\beta}$, es la siguiente:

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}. \quad (3.19)$$

Estos coeficientes se pueden estimar mediante el Estimador de Mínimos Cuadrados siempre y cuando exista independencia entre las variables. Cuando los términos están muy correlados, las columnas de la matriz X tendrán una dependencia lineal entre ellas y, por tanto, la matriz es casi singular y el estimador de mínimos cuadrados se convierte muy sensible a errores aleatorios en la variable objetivo, produciendo mucha varianza en los coeficientes.

Además, para que este método de predicción funcione se debe asumir que los datos se pueden obtener como una combinación lineal de los x_i y, por ello, los resultados no suelen ser demasiado precisos. Pese a esto es un método que sirve como línea base para el resto de modelos.

La Regresión Lineal, pese a su sencillez y limitaciones, sigue siendo un método usado de forma amplia en la literatura aunque suele ir acompañado de otras técnicas más sofisticadas. Muchas de estas aplicaciones hacen uso de precios pasados como regresores más otras variables lo que podría ser visto como métodos de series temporales autorregresivos con variables exógenas como hemos visto en la Sección 3.1. En otras referencias como en [30] se usa una regresión múltiple para predicciones a medio-largo plazo en el mercado de electricidad de Zhejiang, provincia de China.

3.2.2. Regresión Ridge

Anteriormente, hemos dicho que la Regresión Lineal tradicional tiene problemas cuando no existe independencia entre las variables. En concreto, cuando existe colinealidad, es decir, cuando para dos variables x_1 y x_2 ocurre que $c_1x_1 + c_2x_2 \simeq c_0$ para ciertas constantes c_1 , c_2 y c_0 es un caso particular en el que la Regresión Lineal no funciona de manera correcta. Para ello, o bien se eliminan de alguna manera esas colinealidades con, por ejemplo, el método de Análisis de Componentes Principales, técnica de reducción de la dimensionalidad que genera variables ortogonales [28] o bien un recurso muy utilizado es el uso de la Regresión Ridge, propuesto por Hoerl y Kennard en 1970 [29].

Matemáticamente la Regresión Ridge, consiste en reemplazar el sistema de ecuaciones lineales descrito para la Regresión Lineal habitual por el problema de encontrar unos β que minimicen:

$$\|X\beta - \mathbf{y}\|^2 + \alpha\|\beta\|^2 \quad (3.20)$$

dado un factor $\alpha > 0$ elegido apropiadamente. Una solución explícita, denotada $\hat{\beta}$, es la siguiente:

$$\hat{\beta} = (X^T X + \alpha I)^{-1} X^T \mathbf{y} \quad (3.21)$$

donde I es la matriz identidad $(n + 1) \times (n + 1)$. El valor de α de alguna manera regula la importancia entre el peso del término de regularización y la suma de los errores al cuadrado, función a minimizar en la Regresión Lineal. Para $\alpha = 0$, esto se reduce al método de mínimos cuadrados, siempre que $(X^T X)^{-1}$ exista.

3.2.3. Árboles de decisión

Los árboles de decisión son un método supervisado no paramétrico que puede ser usado tanto para clasificación como para regresión. Es un método muy usado y descrito en profundidad en diferentes referencias como, por ejemplo, en [15]. El objetivo es crear un modelo que prediga el valor de la variable objetivo aprendiendo reglas básicas inferidas a partir de las variables de los datos y que definen regiones cuyas aristas son siempre paralelas a los ejes. Dentro de cada región se asigna una función sencilla, en ocasiones, una constante.

Los árboles de decisión, al contrario que otros métodos de predicción, pueden recibir tanto variables categóricas como variables numéricas como entrada. Las reglas con variables numéricas son del estilo $x_1 > \theta_1$ mientras que las reglas con variables categóricas suelen ser del estilo $x_1 = \theta_1$.

Para que el modelo aprenda a partir de los datos de *train*, tenemos que determinar la estructura del árbol, incluyendo qué variables habrá en cada nodo, así como el valor del parámetro θ que servirá para definir las distintas regiones. Por último, debemos determinar qué valor o valores tomarán los elementos en una determinada región y la profundidad que tendrá el árbol.

Consideremos únicamente el problema de regresión que es el que atañe a la predicción de precios. Elegir las regiones generadas por el árbol mediante la minimización del error cuadrático medio sería prácticamente imposible en la gran mayoría de los casos por la gran variedad combinatoria que tendría el problema. Por tanto, generalmente se usan técnicas de optimización avariciosa comenzando en un solo nodo raíz correspondiente a todo el espacio de entrada y haciendo crecer el árbol añadiendo nodos uno a uno. En cada iteración habrá una serie de regiones candidatas para ser divididas, correspondiente a la adición de un par de nodos hoja en

esa región en el árbol existente. La elección de tanto la región y la variable como del criterio de división puede ser realizada de forma eficiente con una búsqueda exhaustiva notando que dada una región de división de una variable y su límite, la elección óptima está dada por la media local de los datos. Este proceso se repite para todas las posibles elecciones de la variable a dividir y la que resulte con el menor error cuadrático medio es la elegida.

Después de describir el proceso de elección de la estructura del árbol aún nos falta saber cómo conseguir cuándo parar de dividir nuestro conjunto de datos. Esto puede ser realizado estableciendo una tolerancia a partir de la cual ya no es necesario dividir más los nodos, o definir un número máximo de registros en cada nodo hoja. Habitualmente se establecen dos tipos de estrategias: o bien se poda el árbol de forma previa estableciendo estos criterios o se hace a posteriori de manera que se permite que el árbol crezca hasta dejar cada registro del conjunto de datos en un nodo hoja. Todos esos parámetros pueden y deben ser optimizados para obtener el mejor modelo. Los árboles de decisión tienen numerosas ventajas [39] por las cuales siguen siendo modelos de referencia:

- Son simples de entender y de interpretar ya que los árboles pueden ser visualizados. Además dada una situación observable en el modelo, la explicación para la condición es fácilmente expresable mediante lógica booleana. Por ello, son ampliamente usados en caso de necesitar algoritmos que necesiten una posterior interpretación de los resultados.
- Necesitan poco preprocesamiento de los datos ya que no es obligatorio utilizar una normalización de las variables de entrada, pueden trabajar con datos no numéricos, etc.
- El coste de uso de un árbol de decisión en predicción es de media logarítmico en el número de puntos usados para construir el árbol puesto que ese es el coste de búsqueda en un árbol binario.
- Se puede validar el modelo usando tests estadísticos de manera que se hace posible establecer una confianza estadística para el modelo resultante.

Por contra, sus inconvenientes son:

- Los cortes sólo pueden ser paralelos a los ejes del espacio de atributos [15].
- Debido al punto anterior, los árboles de decisión a veces necesitan crear modelos demasiado complejos, con demasiada profundidad y muchas reglas, para mejorar la precisión en la fase de entrenamiento pero que luego no generalizan bien. Algunos mecanismos usados para evitar este problema, también llamado **sobreentrenamiento**, son realizar podas en el árbol, establecer un mínimo de ejemplos en los nodos hoja o establecer una profundidad máxima. Este problema también suele ser resuelto mediante el uso de *ensembles* como el modelo de *Random Forest*.
- El problema de encontrar el árbol de decisión óptimo es NP-completo [33]. En la práctica los árboles de decisión están basados en algoritmos heurísticos que llegan a soluciones localmente óptimas y, por tanto, no garantizan una solución globalmente óptima. Este problema suele ser mitigado de nuevo con *ensembles*, donde los atributos y las muestras son elegidos aleatoriamente encontrando distintas soluciones locales.

Este método es usado en [37] junto con una red neuronal para predecir en cada una de las regiones en las que el árbol de decisión divide el conjunto, con el objetivo de minimizar el error que se produce usando la media del precio en cada una de las regiones.

3.2.4. K-Vecinos Próximos

El método de los **K-Vecinos Próximos (K-NN)** se basa en inferir la variable a predecir usando los K casos en el conjunto de entrenamiento que sean más parecidos al nuevo dato. El número K de vecinos a usar en el entrenamiento pueden ser definidos por el usuario y, por tanto, deberá ser hiperparametrizado puesto que cambiando el número de vecinos a usar puede mejorar o empeorar los resultados del algoritmo. Otra forma de definir el número de vecinos es usar un radio de distancia en el que los vecinos que se encuentren dentro de ese radio son usados para predecir la variable objetivo. Aunque, en general, se puede usar cualquier distancia como, por ejemplo, la distancia euclídea, uno de los aspectos más importantes y difíciles de resolver de este método es conseguir una función de distancia eficaz.

Otro aspecto a tener en cuenta es la importancia que va adquirir cada uno de los vecinos, ya que podemos decidir que el peso sea uniforme y, por tanto, para el caso de regresión usar la media aritmética de los vecinos, o por el contrario que los pesos vayan en función de la distancia a la que se encuentran y, por tanto, la predicción sea la media ponderada de los vecinos.

Existen varios algoritmos para calcular los vecinos próximos de los cuales destacan los siguientes [1]:

- **Fuerza bruta.** Es la implementación más simple posible en la que se realiza el cómputo de todas las distancias entre los puntos del dataset: para N registros con D variables, esta aproximación tiene un coste $O[DN^2]$. Por tanto, para pocos datos es suficientemente eficiente pero cuando el número de registros en el dataset crece, este método comienza a ser imposible de calcular rápidamente.
- **K -D Tree.** [14] Para mejorar las ineficiencias computacionales del algoritmo por fuerza bruta, existe una variación del algoritmo en la que se usan estructuras basadas en árboles. Este algoritmo intenta reducir el número requerido de distancias que necesitan ser calculadas. La idea básica es que si el punto A es muy distante del punto B y el punto B es cercano a C entonces A y C serán muy distantes y no será necesario calcular la distancia entre A y C de manera explícita. De esta manera, el coste computacional se reduce hasta conseguir un coste de $O[DN \log(N)]$.
- **Ball Tree.** [38] *Ball Tree* particiona los datos en los ejes cartesianos mientras que la partición de este algoritmo son series de hiperesferas anidadas. Esta construcción es más costosa que la de *K-D Tree*, pero los resultados son mejores para algunos tipos de datos.

El uso de estas técnicas para el problema de predicción de precios ha servido para resolver el problema de precios extremadamente altos o bajos y reemplazar los picos de precios que se consideran anormales por los de otros días similares a ese como en [49]. También han sido propuestos para la predicción de los precios directamente como en [35] que hace uso de un algoritmo K-NN con pesos basados en las distancias euclídeas entre los vecinos para predecir días laborables únicamente. Además en esta publicación estos pesos son estimados haciendo uso de algoritmos genéticos.

3.2.5. Máquinas de Vectores Soporte

Las Máquinas de Vectores Soporte, en inglés **Support Vector Machines (SVM)**, son algoritmos de aprendizaje supervisado que han tenido una gran repercusión en los últimos

años después de que Vapnik publicase su libro en la materia [46]. Este modelo de aprendizaje automático puede ser usado tanto para problemas de regresión como para clasificación. Las máquinas de vectores soporte determinan los parámetros del modelo mediante la resolución de un problema de optimización convexa.

En primer lugar la SVM fue propuesta para resolver problemas de clasificación de dos clases. La idea básica para estos problemas de clasificación de dos clases consiste en construir un hiperplano que separe de forma óptima a los puntos de una clase de la otra, es decir, buscan el hiperplano que tenga la máxima distancia con los puntos que estén más cerca de él mismo. Por eso también a veces se les conoce a las SVM como clasificadores de margen máximo. Este hiperplano ejerce de frontera de decisión entre las dos clases; de esta forma, los puntos del vector que son etiquetados con una categoría estarán a un lado del hiperplano y los casos que se encuentren en la otra categoría estarán al otro lado.

Para el primer problema para el que fue propuesto este modelo, problemas de clasificación con dos clases, podemos definir el conjunto de datos formalmente de la siguiente manera:

$$D = \{(x_i, y_i) : x_i \in \mathbb{R}_p, y_i \in \{-1, 1\}\}_{i=1}^n. \quad (3.22)$$

Si suponemos que el conjunto de datos es separable, entonces se puede definir el hiperplano que separa las dos clases de la siguiente manera:

$$wx + b = 0. \quad (3.23)$$

Por supuesto, pueden existir muchos hiperplanos, incluso infinitos, que separen la muestra en las dos clases. Es aquí donde aparece uno de los conceptos centrales en las SVM, el concepto de margen, que se define como la distancia mínima entre el hiperplano que ejerce como frontera de decisión y cualquiera de las muestras. En las SVM se obliga a que este margen sea máximo y, por tanto, queremos optimizar w y b para maximizar la distancia entre los puntos y la frontera. Finalmente, se puede demostrar que esto es equivalente a resolver el siguiente problema de optimización:

$$\arg \min_{w,b} \frac{1}{2} \|w\|^2 \quad (3.24)$$

sujeto a las siguientes restricciones:

$$\begin{aligned} wx_i + b &\geq 1 && \text{para } x_i \text{ de la clase 1,} \\ wx_i + b &\leq -1 && \text{para } x_i \text{ de la clase -1.} \end{aligned} \quad (3.25)$$

Haciendo uso del Lagrangiano y de la representación dual del problema de maximización del margen finalmente obtenemos que el hiperplano óptimo es:

$$f(x) = \sum_{i=1}^N a_i y_i x_i + b. \quad (3.26)$$

En el caso de que el conjunto de datos no sea separable, se introduce una nueva variable, llamada holgura y denotada como ξ que permite ciertos errores de clasificación. A la vez, ahora se vuelve necesario penalizar esos errores cometidos. Para ello se introduce un nuevo parámetro

C que cuanto mayor es más se penalizan los errores y, por tanto, se permiten menos. El control de este parámetro es importante puesto que cuanto mayor sea el valor de C , más probable es que se produzca *overfitting* y viceversa.

Por tanto, ahora el objetivo se convierte en maximizar el margen al mismo tiempo que se penalizan los puntos que se encuentran en el lado incorrecto de la frontera de decisión y, por tanto, debemos minimizar:

$$C \sum_{i=1}^N \xi_i + \frac{1}{2} \|w\|^2. \quad (3.27)$$

En el caso de usar SVM para problemas de regresión, como la predicción de precios, se suele llamar SVR y se define de forma similar a como se hace en la Regresión Lineal pero en lugar de usar una función de error cuadrática, esta función se reemplaza por una función de error E_ϵ que devuelve 0 si la diferencia entre la predicción y la variable objetivo es menor que ϵ y un error, como, por ejemplo, el valor absoluto entre predicción y real, en caso contrario. Con este planteamiento la función de error a minimizar es:

$$C \sum_{i=1}^N E_\epsilon(f(x_i) - y_i) + \frac{1}{2} \|w\|^2 \quad (3.28)$$

De nuevo, haciendo uso de Lagrange y de la forma dual y operando, se puede obtener la siguiente función de decisión:

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + \rho \quad (3.29)$$

Las SVM son usadas para la predicción de precios en una importante variedad de publicaciones: en [42] que hace una comparativa entre Redes Neuronales y SVM, en [53] que las usa para realizar predicciones precisas y también para aportar un intervalo de confianza. Existen varias propuestas de modelos híbridos usando SVMs como en [20] que realiza un modelo híbrido con SVM para capturar los patrones no lineales y ARIMA; este modelo híbrido se denomina en esta publicación SVRARIMA.

3.2.6. Redes Neuronales Artificiales

Las **Redes Neuronales Artificiales** son un tipo de algoritmo de Aprendizaje Automático que se inspira en el funcionamiento neuronal de los seres vivos. Una Red Neuronal contiene varias unidades de procesamiento que se conectan entre sí formando distintas arquitecturas. Cada unidad o neurona artificial, simula el funcionamiento de una neurona: se activa si la cantidad total de señal que recibe excede su umbral de activación. En este caso, el nodo se activa y emite una señal al resto de las neuronas adyacentes. Por tanto, cada unidad se convierte en un transmisor de la señal que puede aumentar o disminuir dicha señal.

La primera Red Neuronal fue propuesta por Frank Rosenblatt en 1958 [41]. Esta red consistía en un Perceptrón de una sola capa, es decir, la memoria de la neurona es una serie de pesos de entrada en forma de vector, $W = (w_1, w_2, \dots, w_n)$, que se aplica sobre el vector de entrada $\mathbf{x} = x_1, x_2, \dots, x_n$ obteniendo una suma ponderada de los valores de entrada. Por último la función de activación devuelve 1 si el resultado de la operación anterior es mayor que un cierto umbral

y -1 en caso contrario.

El proceso de aprendizaje es iterativo: en cada época de aprendizaje el modelo modifica los pesos para minimizar una función de coste que depende del problema en particular.

Las redes de una sola capa tienen la desventaja de que sólo pueden resolver problemas lineales. Esto provocó la aparición de las redes multicapa. El primer algoritmo de entrenamiento para redes multicapa fue desarrollado por Paul Werbos en 1974 [47].

En el caso del **Perceptrón Multicapa**, también llamado *feed-forward neural network*, la arquitectura se complica ya que se introduce al menos una de las llamadas capas ocultas. La forma que tienen de aprender estos modelos es mediante el llamado algoritmo de *back propagation* que compara la salida obtenida con la real de manera que va modificando los pesos en las distintas conexiones para converger a la salida objetivo.

Matemáticamente podemos describir las redes neuronales de la siguiente manera. Primero definiremos el proceso de cómo la información fluye hacia adelante y posteriormente el paso de la propagación hacia atrás del error. Por tanto, primero se construyen M combinaciones lineales de las variables de entrada x_1, \dots, x_D de la siguiente manera

$$a_j = \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \quad (3.30)$$

donde $j = 1, \dots, M$ es el número total de salidas, y el superíndice (1) indica que los correspondientes parámetros se encuentran en la primera capa de la red.

Las cantidades a_j se conocen como activaciones. Cada una de ellas es transformada usando una función de activación no lineal y diferenciable $h(x)$ obteniendo así

$$z_j = h(a_j). \quad (3.31)$$

Estas cantidades corresponden a las salidas de las llamadas capas ocultas. Las funciones no lineales suelen ser escogidas como funciones sigmoideas. Estos valores suelen ser linealmente combinados de nuevo para obtener nuevas salidas de unidades de activación.

$$a_k = \sum_{j=1}^M w_{kj}^{(2)} z_j + w_{k0}^{(2)} \quad (3.32)$$

donde de forma similar, $k = 1, \dots, K$, es el número total de salidas, y el superíndice (2) indica que los correspondientes parámetros se encuentran en la segunda capa de la red. La función de activación final depende del problema a resolver y en el caso de regresión se usa la identidad $y_k = a_k$.

En la fase de entrenamiento el objetivo es optimizar los parámetros \mathbf{w} de manera que se minimice el error cometido $E(\mathbf{w})$. El error $E(\mathbf{w})$ va a alcanzar su mínimo cuando $\nabla E(\mathbf{w}) = 0$. El problema es que típicamente el error tiene una fuerte componente no lineal con los pesos \mathbf{w} y, por tanto, van a existir muchos puntos que cumplan que el gradiente es 0 y que no sean mínimos globales. Por todo esto, el objetivo es encontrar una heurística que al menos alcance un mínimo local suficientemente bueno. Podemos considerar el error como la suma de los errores de cada

uno de los ejemplos del conjunto de datos. Es decir,

$$E(\mathbf{w}) = \sum_1^N E_n(\mathbf{w}). \quad (3.33)$$

Para cada patrón del conjunto de entrenamiento se realizan las operaciones de propagación hacia adelante. Si realizamos la derivada de E_n respecto al peso w_{ji} para cada unidad tenemos, aplicando la regla de la cadena

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{\partial E_n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} = \frac{\partial E_n}{\partial a_j} z_i. \quad (3.34)$$

Para las unidades de salida se tiene

$$\frac{\partial E_n}{\partial a_j} = y_k - t_k \quad (3.35)$$

donde t_k es la variable objetivo y y_k la predicción realizada. En las unidades ocultas tenemos

$$\frac{\partial E_n}{\partial a_j} = \sum_k \frac{\partial E_n}{\partial a_k} \frac{\partial a_k}{\partial a_j} \quad (3.36)$$

donde la suma recorre todas las unidades k cada una de las cuales tienen j conexiones. La fórmula anterior se puede reescribir para obtener la llamada fórmula de *backpropagation*:

$$\frac{\partial E_n}{\partial a_j} = h'(a_j) \sum_k w_{kj} \frac{\partial E_n}{\partial a_k}. \quad (3.37)$$

Las Redes Neuronales han sido ampliamente usadas para el problema de predicción de precios con distintas variables de entrada, distintos preprocesamientos de los datos, eliminación de *outliers*, etc. Existen algunas publicaciones en las que hay una sola unidad de salida para predecir la siguiente hora, la hora en el horizonte h , el pico, etc; existen otros grupos de publicaciones que tienen 24 o 48 unidades de salida. En concreto, los más relevantes para este estudio son:

- [45], que es uno de los primeros estudios sobre el tema y en el que se usa una arquitectura con 15 parámetros de entrada, 15 unidades ocultas y 1 salida para predecir el precio para el *Victorian Power System*
- [24] donde se utilizan datos del mercado español para realizar una comparativa entre varios modelos, en concreto, uno de los propuestos es un perceptrón multicapa con una arquitectura de una capa oculta y haciendo uso del viento y la demanda como variables regresoras
- También es relevante [27] donde usan una combinación de Redes para predecir el valor máximo, mínimo y medio que es proporcionado a 5 redes neuronales principales para predecir el precio finalmente.
- En [21] usan redes neuronales de una sola capa pero utilizando un algoritmo distinto al *backpropagation*, este algoritmo se denomina *extreme learning machine (ELM)* mediante el cual los parámetros son seleccionados de manera aleatoria o analíticamente en lugar de tener que ser ajustados de forma iterativa.

3.2.7. *Ensemble Learning*

Un *ensemble* es un conjunto de modelos predictivos que son combinados de forma ponderada para producir una predicción única con, normalmente, mejores resultados que los modelos individuales por separado. Estos métodos suelen ser efectivos debido a que hacen uso de la idea de que cuando los modelos individuales toman decisiones contrapuestas, se espera que la mayoría produzcan una predicción más precisa de manera que al ponderar todas las predicciones individuales se equilibren y cancelen los fallos individuales, de manera que la conjunta consiga un acierto mucho mayor. Es por ello que normalmente vamos a querer predictores que sean bastante distintos entre sí puesto que si todos predicen resultados similares no habrá contraposición.

Existen muchas técnicas para la construcción de *ensembles*, sobre todo dos muy famosas basadas en árboles: *boosting* [43] y *bagging* [17]. En *boosting*, los sucesivos árboles dan un peso extra a los puntos incorrectamente predichos por los predictores previos y la predicción final se realiza mediante una votación; en *bagging* los sucesivos árboles no dependen en los árboles anteriores que son creados usando *bootstrap*, es decir, con un muestreo aleatorio con reemplazamiento; al final se realiza una votación simple entre todos. Por último, en este trabajo nos vamos a centrar en el algoritmo *Random Forest* propuesto por Breiman en 2001 [18] y que será detallado en la siguiente subsección.

Random Forest

El algoritmo *Random Forest* es un *ensemble* muy efectivo y muy utilizado en la actualidad. Cada árbol, al igual que en *bagging*, se genera mediante un muestreo aleatorio con reemplazamiento del conjunto de *train* original. *Random Forest* difiere en la manera que están contruidos los árboles. El algoritmo para Random Forest puede ser escrito de la siguiente manera [34]:

1. Se eligen n submuestras de los datos originales mediante *bootstrap*.
2. Para cada una de las submuestras se construye un árbol de decisión sin podar pero que en cada nodo en lugar de elegir el mejor corte posible entre todos los atributos, se eligen aleatoriamente un subconjunto de ellos.
3. Se predicen los nuevos datos realizando una agregación de todas las predicciones individuales mediante una media en el caso de regresión.

Debido a que con *Random Forest* se genera una mayor aleatoriedad, la diversidad en estos tipos de *ensembles* es superior a otros, que es justo lo interesante cuando tratamos con *ensembles*, puesto que necesitan modelos que ofrezcan predicciones muy distintas. Estos modelos proporcionan mejores resultados empíricos en muchos conjuntos de datos y sus ventajas son:

- Se generan modelos predictivos robustos frente a *outliers* y ruido.
- Es más rápido que *bagging* o *boosting* debido a que en cada decisión de división se tienen menos variables entre las que decidir.
- Puede proporcionar información sobre la relevancia de las variables: si una variable es muy elegida entre los árboles para la predicción es que es una buena variable predictora; por el contrario, si nunca es escogida es porque tiene poco poder predictivo.

- Es fácilmente paralelizable debido a la generación de muchos modelos al mismo tiempo.
- Proporciona una cierta idea de proximidad, elementos que caen en los mismos nodos terminales a menudo son nodos similares y viceversa.

Con este método no existen muchas referencias y la mayoría son muy actuales; en concreto destaca [36] que usa los *Random Forest* para la predicción del precio en el Mercado Eléctrico de Nueva York y usa como predictores la propia serie temporal de precios desfasada 3, 24, 168 y 720 horas, la demanda, la temperatura y un indicador de día de la semana.

3.3. Variables relevantes

Las variables relevantes para el problema de predicción de precios más usadas en la bibliografía [13] se pueden dividir en las siguientes categorías:

- **Generación de renovables.** Como vimos en el Capítulo 2, las renovables participan en el mercado a precio 0 y, por tanto, abaratan los precios. Por ello, es importante tenerlas en cuenta a la hora de predecir el precio. Ahora bien, no tendremos el dato real para el día para el que tenemos que dar la predicción. Por ello, es necesario realizar predicciones de producción que serán las que entren como variables al modelo. Las dos tecnologías de generación de renovables más importantes en España y que, por tanto, deben ser tenidas en cuenta como variables de entrada para el modelo son:
 - Predicción de producción eólica.
 - Predicción de producción solar (fotovoltaica y solar).
- **Demanda.** Al igual que la producción de las renovables, la demanda es una variable que afecta de forma directa a los precios de la energía ya que cuanto más demanda existe, las ofertas de compra son mayores y, por tanto, el precio sube.
- **Exportaciones/Importaciones.** Como también mencionamos en la descripción del funcionamiento del mercado eléctrico, se pueden producir exportaciones e importaciones que afectan al precio.
- **Precio en otros países.** Las importaciones y exportaciones se producen debido al precio de la electricidad en otros países. Es por ello que el precio de otros países, en concreto, el precio francés, que es el principal exportador de energía eléctrica a España, afecta al precio español. Por este mismo motivo, como veremos más adelante, existe una gran correlación entre el precio francés y el español.
- **Variables meteorológicas.** Las variables meteorológicas afectan tanto a las producciones de renovables como a la demanda y, por tanto, serán necesarias para las predicciones de ambas variables pero también tienen relación directa con el precio de la electricidad. Las más usadas en la bibliografía y que han sido tenidas en cuenta son:
 - Temperatura
 - Viento
 - Hidricidad
 - Radiación

- **Variables económicas.** Existen variables económicas que afectan al precio de la energía porque se encuentran relacionadas con la demanda y con la producción de energías no renovables como pueden ser el PIB, el precio del gas y el petróleo, etc. La dificultad de encontrar estas variables para el periodo de estudio necesario y con una granularidad suficiente han conllevado que sean dejadas para trabajo futuro.
- **Tipo de día y hora.** Como hemos visto en el Capítulo 2, el precio depende del tipo de día al que nos enfrentemos. Esto se puede intuir con facilidad, debido a que los hábitos de consumo eléctrico no son iguales en invierno que en verano, ni a unas horas o a otras, ni un día de la semana u otro. Es por ello que todo esto debe ser tenido en cuenta a la hora de crear el modelo. En este estudio se usarán las siguientes variables:
 - Hora.
 - Día de la semana.
 - Día del mes.
 - Mes.

4

Propuesta de servicio de predicción

“It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.”

Arthur Conan Doyle, Sherlock Holmes

En este capítulo vamos a abordar el principal problema que conlleva este estudio, que no es otro que la consecución de un modelo eficaz para el problema de la predicción de precios haciendo uso de datos reales del mercado *spot* español.

Se ha seguido la metodología habitual en problemas de ciencia de datos y aprendizaje automático:

- Definición de las variables que sean relevantes para el problema de predicción de precios.
- Análisis de las variables que serán usadas como datos de entrada al modelo para hacer una primera selección de variables.
- Normalización de la variable objetivo con el fin de reducir la varianza.
- Selección de los modelos que harán uso de esas variables para establecer las distintas predicciones.
- Hiperparametrización de las variables de los modelos teniendo en cuenta las distintas restricciones y buenas prácticas de cada uno.
- Con todos los modelos resultantes de la hiperparametrización se seleccionan los que minimicen las métricas de error elegidas como más adecuadas para el problema.
- Realización de un análisis exhaustivo de los errores del mejor modelo resultante.

4.1. Datos para los experimentos

Los datos para la predicción de precios del mercado spot español con los que se ha llevado a cabo la prueba así como otros datos relevantes para la predicción como el precio de la electricidad en Francia, la demanda o la generación de renovables han sido obtenidos de la página web de e-SIOS de REE [5] donde existen datos a partir del 1 de enero de 2014 hasta la actualidad. Esta web permite descargar los datos de forma manual pero también tiene una API para el acceso a los mismos datos que se pueden descargar manualmente. Esto permite que las simulaciones que se han llevado a cabo en este estudio se puedan convertir en un producto que descargue los datos necesarios mediante llamadas a la citada API.

Además de los datos de e-SIOS se dispone del coeficiente de festividad (CF) descrito en Ecuación 2.1, obtenido como un derivado de los datos publicados en el BOE sobre los festivos en España.

Por último, existe otra fuente de datos meteorológica proporcionada por el IIC mediante la cual se han obtenido temperaturas horarias para las capitales de provincia españolas. Estos datos han sido modificados para su uso en la predicción de precios y será explicado con detalle en Subsección 4.2.3

4.2. Análisis de datos y normalización

En la primera parte de nuestra propuesta se ha llevado a cabo un análisis descriptivo de los datos para entender mejor la importancia de todas las variables mencionadas en la Sección 3.3 así como para poder realizar un preprocesado de datos lo más eficaz posible.

4.2.1. Datos estadísticos

Para comenzar, se realiza unas estadísticas descriptivas de las variables, entre las que se encuentran los valores mínimos y máximos, el primer y tercer cuartil, la media y la mediana para cada una de las variables, obteniendo la tabla 4.1. En ella se puede ver, como era de esperar, que los datos de demanda son muy altos en comparación a los demás.

También es remarcable el hecho de que haya precios negativos en Francia, en algunos mercados como el francés o el alemán se permiten valores negativos, al contrario que en el mercado español que su límite inferior es 0, y suelen ocurrir cuando hay una gran cantidad de generación de renovable y un descenso en el consumo de forma considerable. Por tanto, estos datos negativos no son errores en los datos sino que son datos perfectamente válidos y que hay que tener en cuenta.

La última columna de la tabla representa el número de *missing values* o registros para los que no hay datos. En esa columna sólo aparecen los precios 24, 48 y 72 horas antes, porque al realizar el desplazamiento faltan los primeros 24, 48 y 72 datos respectivamente.

Tabla 4.1: Tabla de estadísticas de los valores.

Variable	Min.	1st Qu.	Mediana	Media	3rd Qu.	Max	nº NAs
Demanda	18036	24292	28470	28204	31637	40359	-
Solar	0.00	63.85	592.05	1437.85	2600.03	5835.90	-
Eólica	12.6	2926.0	4909.5	5550.0	7635.0	17430.0	-
PrecioES	0.00	35.13	45.13	44.04	56.01	113.92	-
PrecioFR	-10.69	25.70	35.03	36.62	45.66	874.01	-
PrecioES24	0.00	35.12	45.11	44.02	56.00	113.92	24
PrecioFR24	-10.69	25.69	35.00	36.60	45.60	874.01	24
PrecioES48	0.00	35.10	45.10	44.01	56.00	113.92	48
PrecioFR48	-10.69	25.69	35.00	36.58	45.55	874.01	48
PrecioES72	0.00	35.10	45.10	43.99	56.00	113.92	72
PrecioFR72	-10.69	25.67	35.00	36.57	45.53	874.01	72
CF	0.00	0.00	0.00	0.03	0.00	1.00	-

4.2.2. Detección de *outliers*

Se considera que los *outliers* son aquellos valores X que toman las diferentes variables cumplen alguna de estas condiciones [31]:

$$\begin{aligned} X &< Q1 - 1,5RI, \\ X &> Q3 + 1,5RI \end{aligned} \tag{4.1}$$

siendo $Q1$ el primer cuartil, $Q3$ el tercer cuartil y RI , el rango intercuartílico ($Q3 - Q1$) de la variable en cuestión. Con esta definición, encontramos los *outliers* de la tabla 4.2.

Tabla 4.2: Número de *outliers* por variable.

Variable	Nº Outliers
Demanda	0
Eólica	0
Solar	208
Precio España	487
Precio Francia	391

Si nos fijamos, el número de *outliers*, el precio en España tiene bastantes más que en Francia pero si nos fijamos en la figura 4.1 veremos que los valores atípicos del precio francés se encuentran mucho más lejos de la media. Los datos atípicos, en numerosas ocasiones deben ser tratados para que no empeoren los resultados de las predicciones pero en nuestro caso son datos que son válidos y que queremos que el modelo aprenda de ellos, por tanto, no se va a hacer ningún tratamiento con ellos.

4.2.3. Generación de variables

Además de las variables primarias comentadas anteriormente, se han generado otras variables que pueden ser de gran interés. En el primer grupo de estas variables ya han sido brevemente comentadas anteriormente, son los valores pasados del precio de la electricidad en España y

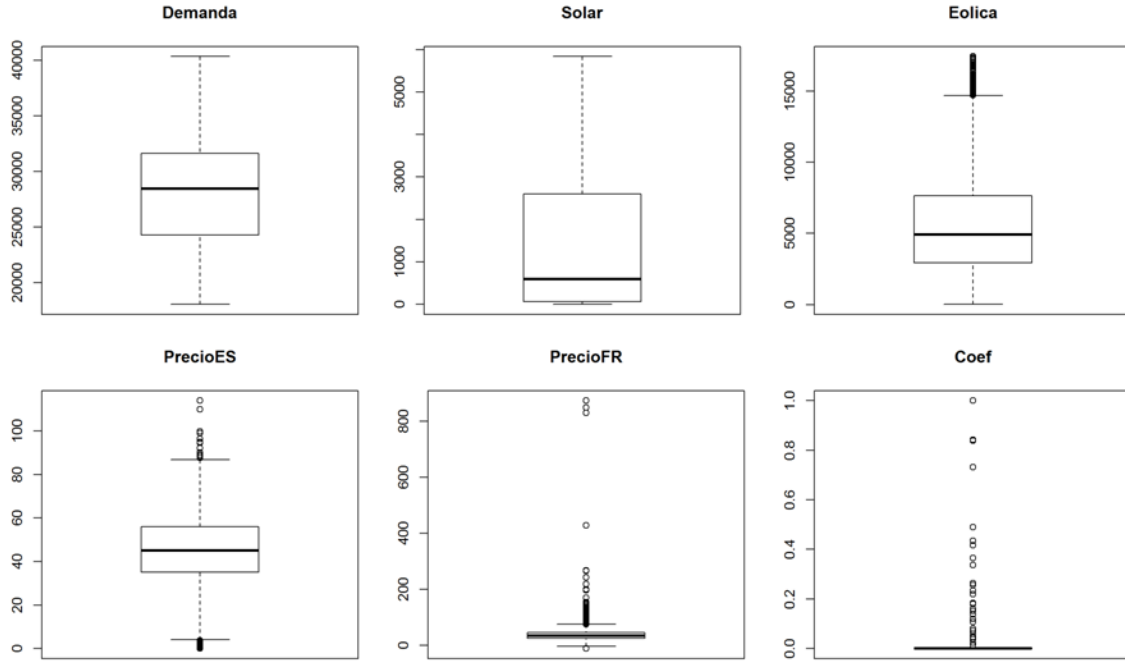


Figura 4.1: Boxplot de las distintas variables. *Elaboración propia.*

Francia. Estas variables representan lo ocurrido a la misma hora uno, dos y tres días antes. Estos valores son muy importantes por la componente estacional horaria que tiene el precio y que fue detallada en la Sección 2.3.

En segundo lugar, existe un último grupo en el que aparecen la variable de la temperatura. Anteriormente se comentó que se disponía de la temperatura en las capitales de provincia. Para contar con un único valor para todo España se ha creado una media ponderada por la población que contiene dicha provincia de manera que la temperatura de provincias como Madrid o Barcelona tendrán más relevancia en la variable que temperaturas en Teruel o Soria.

Además, la temperatura tiene un efecto en el consumo que no es lineal sino que cuando la temperatura supera un cierto umbral se empiezan a usar el aire acondicionado, ventiladores y otros aparatos eléctricos con cada vez más intensidad. Pero también, por otro lado, cuando hace frío se usan calefactores y termos eléctricos para calentar el ambiente de las casas y el agua. Por tanto, en este trabajo se han propuesto las siguientes definiciones de temperaturas:

$$T_2 = |T|, \quad (4.2)$$

que produce el efecto mostrado en la imagen 4.2.

$$T_3 = \begin{cases} 0 & \text{si } T < 10, \\ T & \text{en otro caso} \end{cases} \quad (4.3)$$

que produce el efecto mostrado en la imagen 4.3.

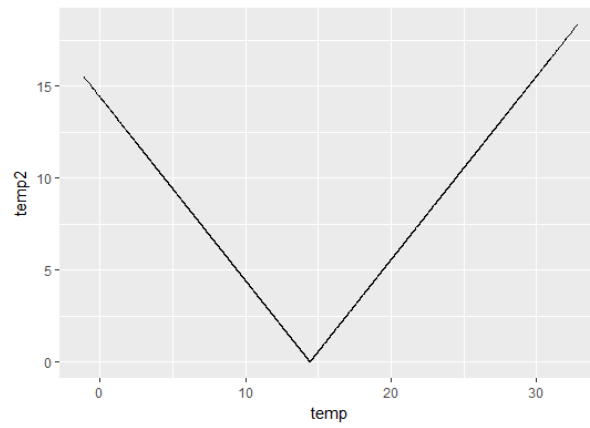


Figura 4.2: Efecto de la temperatura T2 respecto de la original T. *Elaboración propia.*

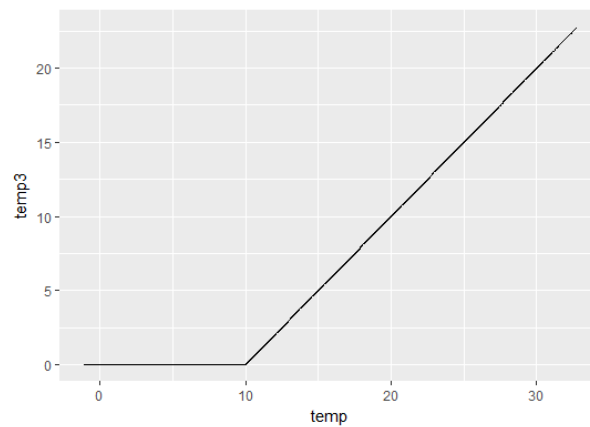


Figura 4.3: Efecto de la temperatura T3 respecto de la original T. *Elaboración propia.*

$$T_4 = \begin{cases} -T & \text{si } T < 15, \\ 0 & \text{si } 15 \leq T < 20, \\ T & \text{en otro caso} \end{cases} \quad (4.4)$$

que produce el efecto mostrado en la imagen 4.4.

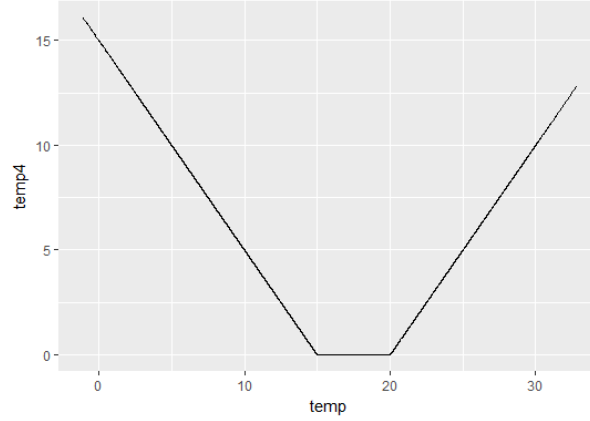


Figura 4.4: Efecto de la temperatura T_4 respecto de la original T . *Elaboración propia.*

$$T_5 = \begin{cases} -T & \text{si } T < 10, \\ 0 & \text{si } 10 \leq T < 15, \\ T & \text{en otro caso} \end{cases} \quad (4.5)$$

que produce el efecto mostrado en la imagen 4.5.

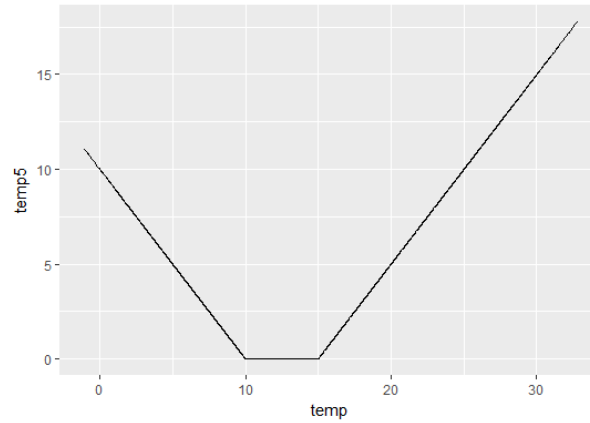


Figura 4.5: Efecto de la temperatura T_5 respecto de la original T . *Elaboración propia.*

$$T_6 = \begin{cases} -T & \text{si } T < 5, \\ 0 & \text{si } 5 \leq T < 20, \\ T & \text{en otro caso} \end{cases} \quad (4.6)$$

que produce el efecto mostrado en la imagen 4.6.

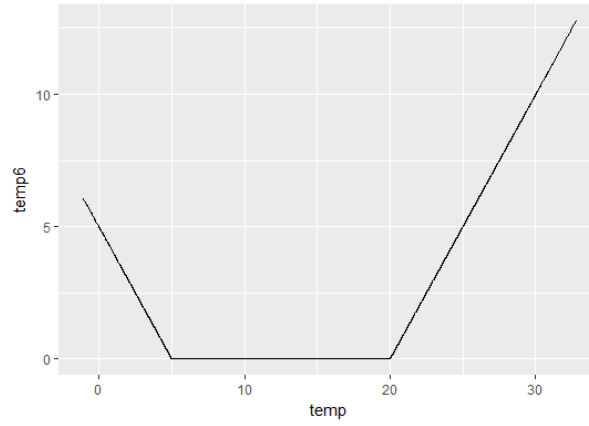


Figura 4.6: Efecto de la temperatura T6 respecto de la original T. *Elaboración propia.*

$$T_7 = \begin{cases} -T & \text{si } T < 5, \\ 0 & \text{si } 5 \leq T < 25, \\ T & \text{en otro caso} \end{cases} \quad (4.7)$$

que produce el efecto mostrado en la imagen 4.7.

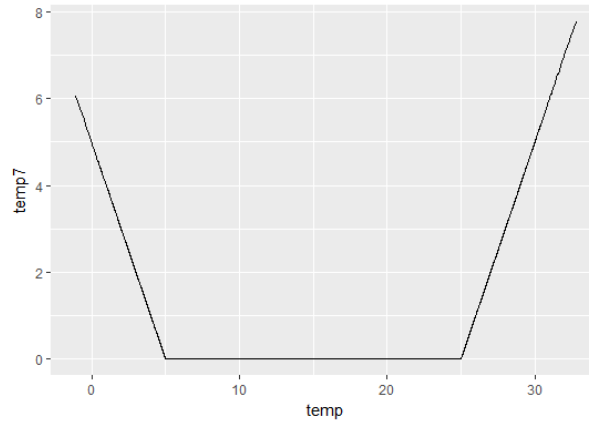


Figura 4.7: Efecto de la temperatura T7 respecto de la original T. *Elaboración propia.*

4.2.4. Análisis de correlaciones

Una de las técnicas más usadas para el análisis descriptivo en este tipo de problemas es el uso de la correlación para conseguir comprender la relación existente entre las distintas variables. Con ese análisis, que puede ser observado en la figura 4.8, podemos destacar la gran importancia de algunas variables que hemos comentado anteriormente. En orden descendente las variables con una mayor correlación son: el precio en España 24, 48 y 72 horas antes, la generación eólica, el precio en Francia y la demanda eléctrica.

Cabe destacar que el precio en Francia tiene bastante relevancia pero es un falso predictor debido a que si queremos predecir el precio en España en el tiempo t no podemos usar el precio

en Francia en ese mismo tiempo t . Ahora bien, pese a que no podemos usar esa variable, sí podemos usar el dato del precio en Francia 24 horas antes y el de 48 horas antes, que siguen siendo relevantes.

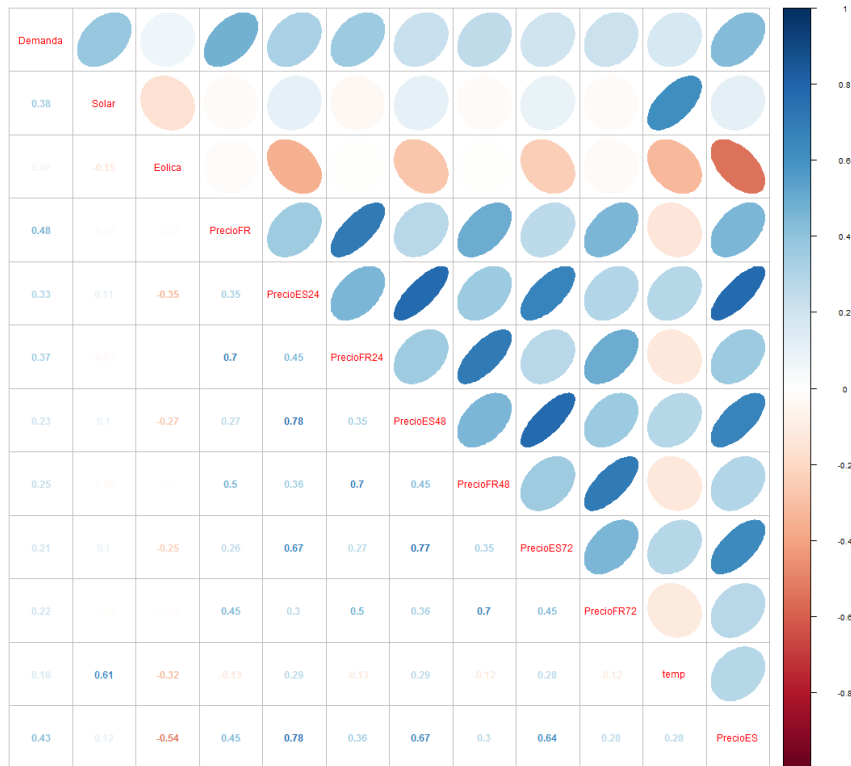


Figura 4.8: Correlaciones de las distintas variables que afectan al precio. *Elaboración propia.*

Además de las variables anteriores, se ha estudiado en detalle la correlación entre las temperaturas. Como podemos ver en 4.9, entre ellas mismas la correlación es bastante alta como es lógico y no tienen mucha correlación los precios. Pese a esto, posteriormente en los modelos se usarán todas ya que el propio modelo las descartará las que no sean lo suficientemente buenas.

4.2.5. Normalización de la variable objetivo

Además del análisis de correlaciones, es muy importante el análisis de la serie temporal de precios ya descrito en la Sección 2.3. Este análisis es probablemente el estudio más importante a la hora de comenzar a construir un modelo predictivo ya que mediante él se consigue comprender el comportamiento de la serie temporal objetivo. Debido a que con el estudio de la serie temporal descubrimos una gran diferencia entre los días laborables y los días no laborables se ha establecido distintos tipos de normalizaciones para conseguir una reducción en la varianza. Estas normalizaciones son funciones que pueden ser invertidas para obtener el valor del precio original. Para el proceso debemos definir dos tipos de normalización: normalización de no laborables y normalización horaria.

Por un lado, la normalización de no laborables conceptualmente pretende eliminar el efecto de disminución en el precio de los días no laborables provocada por la bajada de la actividad laboral. Dentro de ellas se han definido dos tipos de normalizaciones que serán explicadas a continuación.

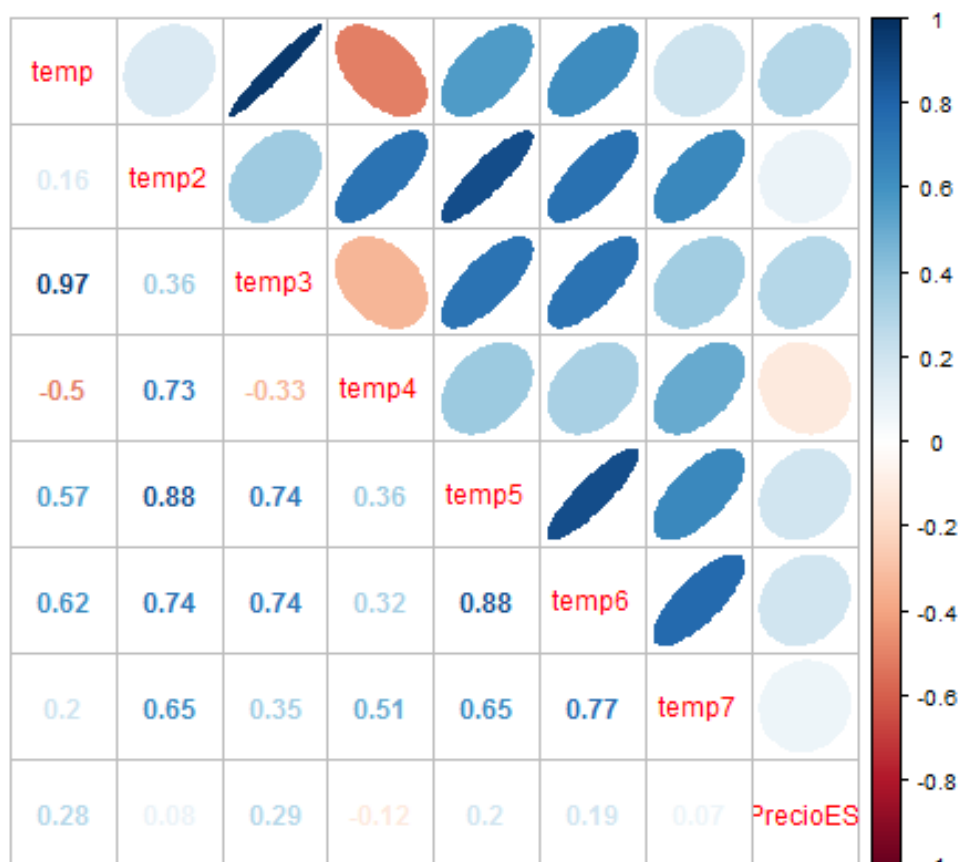


Figura 4.9: Correlaciones de las distintas temperaturas definidas. *Elaboración propia.*

Para ambas necesitamos primero definir las medias horarias de precios por tipo de día del siguiente modo:

$$\bar{p}_{\mathcal{L}h} = \frac{1}{|L|} \sum_{l \in L} p_{lh} \quad \forall h \in [0 - 23], \quad (4.8)$$

donde L es el conjunto de los días que no son sábados, domingos ni festivos,

$$\bar{p}_{\mathcal{S}h} = \frac{1}{|S|} \sum_{s \in S} p_{sh} \quad \forall h \in [0 - 23], \quad (4.9)$$

donde S es el conjunto de los días que son sábados,

$$\bar{p}_{\mathcal{D}h} = \frac{1}{|D|} \sum_{d \in D} p_{dh} \quad \forall h \in [0 - 23], \quad (4.10)$$

donde D es el conjunto de los días que son domingos y

$$\bar{p}_{\mathcal{F}h} = \frac{1}{|F|} \sum_{f \in F} p_{fh} \quad \forall h \in [0 - 23], \quad (4.11)$$

donde F es el conjunto de los días que son festivos en toda España.

En la **normalización de no laborables - aproximación 1**, una vez tenemos las medias se calculan las diferencias entre los días laborables y el resto por hora, es decir:

$$\begin{aligned} dif_{\mathcal{S}h} &= \bar{p}_{\mathcal{L}h} - \bar{p}_{\mathcal{S}h}, \\ dif_{\mathcal{D}h} &= \bar{p}_{\mathcal{L}h} - \bar{p}_{\mathcal{D}h}, \\ dif_{\mathcal{F}h} &= \bar{p}_{\mathcal{L}h} - \bar{p}_{\mathcal{F}h} \end{aligned} \quad (4.12)$$

Por último, se calculan los nuevos precios añadiendo el incremento existente entre los laborables y los que no lo son:

$$\begin{aligned} p'_{sh} &= p_{sh} + dif_{\mathcal{S}h}, \\ p'_{dh} &= p_{dh} + dif_{\mathcal{D}h}, \\ p'_{fh} &= p_{fh} + dif_{\mathcal{F}h} CF_h, \end{aligned} \quad (4.13)$$

donde CF_h es el coeficiente de festividad, definido anteriormente en Sección 2.3, que es un número entre 1 y 0 y significa el porcentaje de la población española para el cual el día dado es festivo.

Por tanto, intuitivamente en esta aproximación lo que estamos haciendo es escalar los precios de los distintos tipos de días no laborables para que se encuentren cerca de la media de los días laborables.

Otra manera de normalizar, nombrada en el presente documento como **normalización de no laborables - aproximación 2**, más usada habitualmente en la bibliografía y que por supuesto también ha sido tomada en cuenta es tomar los nuevos precios como:

$$\begin{aligned} p'_{lh} &= p_{lh} - \bar{p}_{lh}, \\ p'_{sh} &= p_{sh} - \bar{p}_{sh}, \\ p'_{dh} &= p_{dh} - \bar{p}_{dh}, \\ p'_{fh} &= p_{fh} - \bar{p}_{fh} CF_h. \end{aligned} \quad (4.14)$$

En esta otra normalización los precios quedan centrados en el 0.

Por otro lado, la normalización horaria conceptualmente pretende eliminar el efecto horario también descrito en Sección 2.3. Dentro de esta normalización se han definido dos tipos que matemáticamente se definen de forma similar a las anteriores.

- Normalización horaria aproximación 1:

$$p'_h = p_h - dif_h, \quad (4.15)$$

donde en este caso dif_h se define como:

$$dif_h = \bar{p} - \bar{p}_h. \quad (4.16)$$

- Normalización horaria aproximación 2:

$$p'_h = p_h - \bar{p}_h. \quad (4.17)$$

Por tanto, como vamos buscando aquella normalización que minimiza la varianza, vamos a comparar las distintas combinaciones de aproximaciones:

- Datos de precios sin normalizar.
- Usando la normalización de laborables.
- Usando la normalización horaria.
- Usando la normalización de laborables sobre la normalización horaria.
- Usando la normalización horaria sobre la normalización de laborables.

Todas las anteriores combinaciones se llevan a cabo con las dos posibles aproximaciones. Con todas obtenemos el resultado mostrado en la tabla 4.3 donde muchas de las combinaciones mejoran la desviación típica original de los datos que es de 14.89. En dicha tabla, las columnas son aplicadas en primer lugar y posteriormente se aplica el valor en las filas. En concreto la aplicación de la normalización horaria 2 y posteriormente la normalización de laboralidad 2 que elimina primero el efecto horario y después el efecto del tipo de día, reduce la desviación hasta los 13.70, como se ve en la celda resaltada. Como esta normalización es la que minimiza de forma clara las desviaciones típicas de las predicciones es la que vamos a usar como preprocesado de los precios antes de la modelización.

4.3. Elección del modelo

Para la resolución del problema se han usado los siguientes algoritmos de regresión que ya han sido comentados en el Capítulo 3 por ser los más usados tanto para los problemas de predicción de precios como para otros problemas de regresión en general:

- Regresión Lineal.
- K Vecinos Próximos.
- Máquinas de Vector Soporte para Regresión.

- *Random Forest* para Regresión.
- Perceptrón Multicapa.

Tabla 4.3: Tabla de desviaciones típicas según la normalización aplicada

	-	Normalización Horaria 1	Normalización Horaria 2	Normalización Laboralidad 1	Normalización Laboralidad 2
Normalización Horaria 1	14.37	-	-	13.75	15.77
Normalización Horaria 2	14.37	-	-	13.75	15.77
Normalización Laboralidad 1	14.73	13.77	13.77	-	-
Normalización Laboralidad 2	15.74	15.73	13.70	-	-

Para todos ellos se ha seguido una estrategia de hiperparametrización atendiendo a las particularidades de las variables que reciben los distintos modelos. Además se ha realizado una optimización de aquellos parámetros que son comunes a los distintos modelos como son el número de días para el entrenamiento y las funciones de normalización.

Ya ha sido detallada la normalización de la variable objetivo, el precio, pero también es necesario la normalización del resto de variables. Esta normalización se ha realizado mediante la parametrización de la función de normalización. Este proceso es vital para muchos de los algoritmos de *Machine Learning* usados. En concreto, tanto los K-Vecinos Próximos, como en el Perceptrón Multicapa como en las SVM, es necesario aplicar alguna función de normalización que genere variables que se encuentran en los mismos órdenes de magnitud. Explicaremos la problemática con un ejemplo, supongamos que existe una variable que se encuentra entre 0 y 1 y otra variable que se encuentra entre 0 y 100. Una diferencia de 1 en cualquiera de las variables contribuiría de igual manera en estos métodos que tienen en cuenta distancia entre puntos; en cambio, en el primer caso la variación de 1, representa una variación del 10 % con respecto al máximo y en el segundo supone un 100 %.

En el caso de *Random Forest* no es necesario usar ningún tipo de normalización de las variables puesto que *Random Forest* es invariante frente a variación en la magnitud de una variable debido a que no hace ningún tipo de comparativa entre variables sino que va seleccionando los puntos de corte dentro de una sola variable. Por su parte, en la regresión lineal no sería necesaria una normalización pero como estamos usando regresión lineal con penalización, ocurre que sí necesitamos esa normalización.

Por tanto, para todos los modelos en los que es necesario normalizar los atributos de algún modo se ha hiperparametrizado la función de normalización haciendo siempre una búsqueda en rejilla sobre los siguientes cuatro tipos de normalización disponibles en *Scikit learn*:

- **MaxAbs scaler.** Escala cada atributo por el máximo valor que pueda tomar ese atributo.
- **Robust scaler.** Este método usa estadísticos robustos ante *outliers*. Por tanto, resta la mediana y usa el rango intercuartílico para escalar los datos.
- **Standard scaler.** Convierte en 0 la media y escala para obtener varianza unitaria.

- **MinMax scaler.** Se suele usar como alternativa a la anterior y se define matemáticamente del siguiente modo, $z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$.

4.3.1. Regresión Ridge

Como ya se comentó en la Subsección 3.2.2 la Regresión Ridge, es una Regresión Lineal que tiene un parámetro de penalización y que ha sido hiperparametrizado para conseguir el mejor acierto posible.

4.3.2. K-Vecinos Próximos (K-NN)

En el algoritmo de K-Vecinos próximos, se debe definir una métrica y en este caso se ha usado la distancia euclídea. Aparte de la métrica el otro paso más importante en este modelo es la decisión de cuántos vecinos se toman; para ello existen dos algoritmos: en el primero, se usa un radio con el cual se eligen como vecinos aquellos registros, en nuestro problema horas, que se encuentren a una distancia menor o igual que el radio dado; en el segundo algoritmo, simplemente se toman los K vecinos más cercanos.

En los experimentos realizados se ha usado el segundo y, por tanto, el parámetro que ha sido variado para conseguir los mejores resultados ha sido K. A pesar de que se ha elegido este formato, elegir el otro método es equivalente si hiperparametrizamos de manera correcta el radio.

4.3.3. Perceptrón Multicapa (MLP)

En el Perceptrón Multicapa básicamente se ha parametrizado el parámetro *alpha* el cual sirve para la regularización L2 y que ayuda a evitar el *overfitting* penalizando pesos con magnitudes muy grandes. Se ha usado una red neuronal con dos capas de 50 neuronas cada una debido a que, teóricamente, una red neuronal de dos capas con salidas lineales puede aproximar en principio cualquier función continua en un dominio compacto [15].

4.3.4. Support Vector Machine (SVM)

En el caso de las SVM, todas las posibles combinaciones han sido probadas con *kernel* gaussiano y, por tanto, los parámetros más importantes son epsilon, C y gamma, todos ellos optimizables en el paquete *Scikit Learn*. Recordamos que:

- **Gamma** define la anchura de la gaussiana y, por tanto, cuánta influencia ejerce un solo ejemplo en el conjunto de *train*. Valores pequeños de gamma suponen que la influencia de cada ejemplar es pequeña mientras que valores muy altos suponen una gran influencia. Los modelos son muy sensibles a los valores de gamma y, por tanto, si gamma es muy grande, el modelo va a tender a sobreentrenar; por el contrario, cuando es muy pequeño, el modelo es demasiado simplista y no consigue recoger la complejidad de la forma de los datos.
- **C** es el parámetro que pone en una balanza la equivocación que comete el modelo en los ejemplos de *train* contra la simplicidad de la superficie de decisión. Cuando el valor de C es bajo, la superficie de decisión es suave mientras que cuando C es grande tiende a clasificar

todos los ejemplos de forma correcta y por tanto la complejidad de la superficie aumenta considerablemente.

- **Epsilon** determina el nivel de acierto de la función de aproximación. Si epsilon es cero se espera *overfitting* y cuanto mayor sea más simple será el modelo.

4.3.5. *Random Forest*

En el método *Random Forest* se han parametrizado, por un lado, la variable relacionada con la cantidad de árboles que se usan y, por otro, las variables relacionadas con la complejidad de los distintos árboles de decisión que se generan en cada iteración del algoritmo. Por tanto, se han hiperparametrizado los siguientes parámetros dentro de todos los que permite *Scikit Learn*:

- El número de árboles en el "bosque" (**n_estimators**). Cuanto mayor es este parámetro, teóricamente los resultados son mejores aunque también necesita más tiempo de ejecución para computarlos. Además los resultados dejan de ser significativamente mejores a partir de un determinado número de árboles.
- El número de subconjuntos aleatorios de variables a ser consideradas al dividir un nodo (**max_feat**). Cuanto más bajo mayor será la reducción de la varianza aunque también mayor es el *bias*.
- El número mínimo de datos que pueden quedar en cada nodo hoja (**min_samp_leaf**). Cuanto menor sea, más complejos y desarrollados serán los árboles.

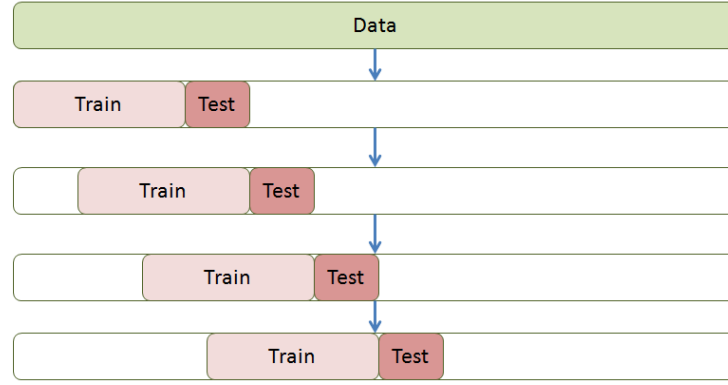
También se tuvieron en cuenta el uso de **min_samp_sp**, que es el mínimo número de datos para dividir un nodo, y **max_depth**, que es el máximo de profundidad al que puede llegar un árbol, pero ambos son dependientes totalmente de **min_samp_leaf** y, por tanto, su hiperparametrización no debería suponer ninguna mejora pero sí producirá que necesitemos un mayor tiempo de ejecución para concluir todas las pruebas.

4.4. Hiperparametrización

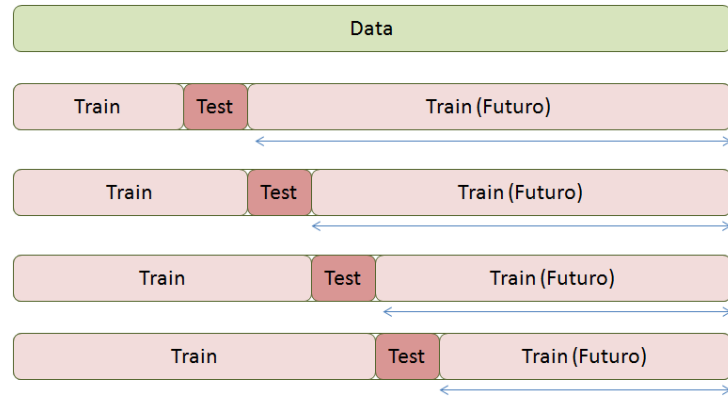
En el proceso de elección del mejor modelo se encuentra, por un lado, la selección del algoritmo de predicción y, por otro lado, nos encontramos con la gran cantidad de posibles parámetros que reciben los distintos algoritmos.

Para llevar a cabo la selección de los mejores parámetros se realiza el proceso de hiperparametrización que básicamente lleva a cabo entrenamientos con un conjunto de datos de *train* dado y predicciones para el conjunto de validación. En este caso debido a que estamos trabajando con series temporales se ha optado por el uso de una validación temporal siguiendo el esquema de la imagen 4.10.

En este esquema se realiza una validación cruzada de forma continua de manera que en ningún caso tomamos valores del futuro para predecir valores del pasado como sí podría ocurrir en una validación cruzada habitual como vemos en la figura 4.11. Esto es necesario puesto que cuando nos encontremos realizando predicciones en un estado real, no podremos usar lo que pasará dentro de varios días.

Figura 4.10: Proceso de validación realizado. *Elaboración propia.*

Asimismo, cabe destacar que el período de *test* siempre es un día completo y el número de días tomados para cada período de *train* ha sido hiperparametrizado, recorriendo de 10 días como mínimo hasta 1 año como máximo, para posteriormente escoger el más adecuado. En general, los resultados muestran que suele afectar el período de *train* elegido, resultando óptimo para la mayoría unos 7-9 meses aproximadamente. Esto es debido porque si no se contempla suficientes casos en el *train* el modelo no tiene información suficiente para generalizar pero tampoco se puede coger un período muy grande hacia atrás porque encontraríamos casuísticas que pueden haberse producido por hechos macroeconómicos que no estamos teniendo en cuenta en este trabajo.

Figura 4.11: Proceso de validación cruzada tradicional. *Elaboración propia.*

Por su parte, el error de validación de cada modelo se tomará como la agregación de los errores producidos para cada día en el año 2016. La explicación de qué error ha sido tomado se realizará en la siguiente sección.

4.5. Métricas de error

En los problemas de *Machine Learning* es muy importante la función de error que tomemos para ser capaces de determinar cuánto aprenden los modelos, realizar una hiperparametrización adecuada y, por tanto, una elección del modelo predictivo y sus parámetros correcta. En otras

palabras es el sistema de evaluación que vamos a usar para los distintos exámenes a los que se ven sometidos los modelos. Esta función de error suele estar definida por el tipo de problema ya que en el caso de problemas de clasificación, como las etiquetas son un número discreto de elementos, se suelen usar métricas como el *accuracy*, *True Positives y Negatives*, *False Positives y Negatives*, Precisión y *Recall*, etc. En el caso de los problemas de regresión, al trabajar con datos continuos, no podemos usar métricas que nos sirvan para medir cuántos aciertos ha realizado el modelo y, por tanto, se usan distancias entre los puntos reales y los predichos de forma análoga, conceptualmente a la figura 4.12. En concreto destacan las dos siguientes métricas:

- Error Absoluto Medio (MAE): mide la media de las diferencias absolutas entre el valor real y el predicho. Matemáticamente:

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} |y_i - \hat{y}_i|. \quad (4.18)$$

- Error Cuadrático Medio (MSE): mide la media de las diferencias al cuadrado entre el valor real y el predicho. Matemáticamente:

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2. \quad (4.19)$$

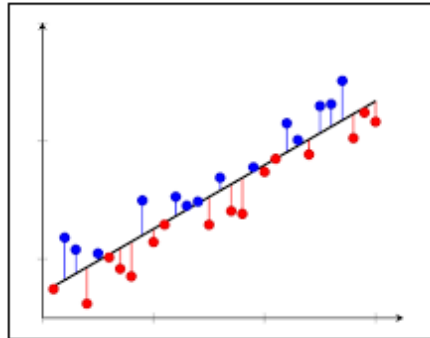


Figura 4.12: Concepto de las métricas de error en regresión. *Fuente [8]*

El MSE es una métrica muy usada por tener ciertas propiedades matemáticas como la diferenciabilidad que no tiene el MAE pero, por contra, tiene un problema para este problema de predicción y es que se enfatiza los errores grandes debido al cuadrado. En nuestro caso los errores grandes se pueden producir cuando el precio es muy alto pero deberían pesar igual que errores más pequeños y, por tanto, se ha decidido tomar como métrica de evaluación el MAE cometido durante todo el período de evaluación, que es el año 2016.

También ha sido usado el error con signo para analizar casos particulares en los que esta métrica permite observar si los errores se distribuyen de forma simétrica y aproximadamente normal al rededor del 0 o si por el contrario se encuentran desplazados por encima o por debajo, lo que supondría que el modelo está sobreprediciendo o infraprediciendo de manera sistemática. Este error se suele denominar Error Medio con Signo (MSD) y matemáticamente se representa del siguiente modo:

$$\text{MSD}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i). \quad (4.20)$$

4.6. Resultados

En esta sección se va a tratar en detalle el resultado de todos los *tests* creados para la predicción del precio en España. Para el *benchmark* se han creado más de 20.000 modelos con los distintos parámetros comentados anteriormente. En concreto, los mejores resultados de cada uno de los modelos son los siguientes:

- En el caso de la **Regresión Ridge**, los parámetros que mejor resultado han obtenido han sido $\alpha = 1,01$, con 90 días para el entrenamiento y normalizando los datos con la normalización *maxAbs*. Con esta configuración vemos que el error absoluto medio se encuentra en torno a 5,75€.

Tabla 4.4: Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de la Regresión Ridge.

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Anual
5.79	7.07	5.40	5.87	6.67	5.49	5.92	5.54	4.65	5.90	4.98	5.75

- Con los **vecinos próximos** encontramos la mejor configuración haciendo uso, de igual modo que la Regresión Ridge, de 270 días para el entrenamiento y la función *robust* para la normalización de los datos y de 20 vecinos. Con estos parámetros se obtiene una reducción del MAE con respecto al valor anterior, obteniendo 5,372€.

Tabla 4.5: Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de K-NN.

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Anual
5.70	7.71	5.35	5.41	5.21	6.19	4.85	4.41	3.49	5.34	5.53	5.37

- En el caso del **Perceptrón Multicapa**, usando los parámetros $\alpha = 1$, la normalización *robust* y como ya se ha dicho antes dos capas ocultas con 50 neuronas se obtiene un error de 5,306€.

Tabla 4.6: Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de MLP.

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Anual
4.73	6.01	5.19	5.19	5.58	4.83	5.23	5.94	4.35	5.40	5.89	5.30

- En el caso de las **Máquinas de Vector Soporte para Regresión**, se alcanza la minimización del error, obteniendo 4,931€ cuando C toma el valor de 100 $\epsilon = 1$, $\gamma = 1$, con normalización *minMax* y con 210 días para el entrenamiento.

Tabla 4.7: Tabla del Error Absoluto Medio por mes y anual para el mejor modelo de SVR

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Anual
5.63	6.60	4.06	4.75	4.77	4.60	4.44	4.00	5.22	4.62	5.69	4.93

- Por último, el modelo que mejor resultado ha ofrecido ha sido el **Random Forest** con $n_estimators = 710$, $min_samp_leaf = 2$, $max_feat = 0.4$ y 270 días de entrenamiento. Con todo ello se ha conseguido 4,44€ de MAE.

Como resultado de todas estas ejecuciones podemos observar que el modelo *Random Forest* es el que ofrece mejores resultados y que en general los meses de verano son más sencillos debido a su menor variabilidad en el precio. El error anual es muy útil para poder comparar entre distintas aproximaciones pero además de esto es necesario un análisis exhaustivo de estos errores para comprender las posibles mejoras que puedan llevarse a cabo en los modelos.

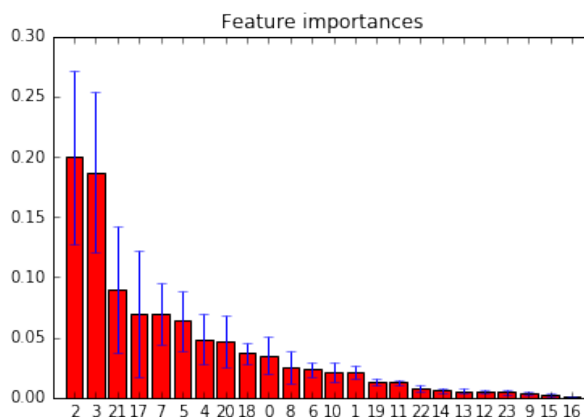
4.6.1. Análisis detallado del mejor modelo

Uno de los apartados más interesantes y que, en particular el mejor modelo, el *Random Forest* nos permite evaluar, es la relevancia de las variables. Esta relevancia se toma como la cantidad de veces que aparece ese atributo en cada uno de los árboles creados. Por tanto, en nuestro caso como hacemos un entrenamiento diario, en cada entrenamiento podemos tener la frecuencia que ha tenido cada variable en ese entrenamiento. A lo largo de todo el año de validación, tenemos 365 frecuencias para cada variable, de manera que si tomamos la media de toda esa serie tenemos la frecuencia media de cada variable. Con estas medias tenemos una buena medida de cómo de importante es cada variable a lo largo del año completo. En las figuras 4.13, 4.15 y 4.18 se numeran las variables de la siguiente forma:

0. Demanda.
1. Producción Solar.
2. Producción Eólica.
3. Precio en España 24h antes.
4. Precio en Francia 24h antes.
5. Precio en España 48h antes.
6. Precio en Francia 48h antes.
7. Precio en España 72h antes.
8. Precio en Francia 72h antes.
9. Coeficiente de festividad.
10. Mes.
11. Día del mes.
12. Día de la semana.
13. Hora.
14. Semana del año.
15. Tipo de día. Que toma los valores de 1 si es sábado, 2 si es domingo, 3 si es festivo y 0 en otro caso.
16. Laborable. Que toma los valores de 0 si es sábado, domingo o festivo y 1 en otro caso.
17. T

18. T_2 19. T_3 20. T_4 21. T_5 22. T_6 23. T_7 **RF con el precio normalizado**

Haciendo uso de la técnica descrita anteriormente para al cálculo de la relevancia de las variables, podemos observar que las más importantes son la energía eólica, el precio en España 24 horas antes y la temperatura en la versión T_5 . Sorprende la baja relevancia que obtiene la demanda cuando en el Capítulo 2 la situábamos como un factor importantísimo debido a que el punto de corte entre demanda y generación se encontrará más arriba o más abajo en función de la cantidad total de energía demandada. Esto es claramente debido a la normalización realizada a la variable objetivo; en esta normalización se elimina el efecto de las horas y de los días de la semana y si no se aplica de igual forma a la demanda provoca que la demanda deje de tener relación con los precios.

Figura 4.13: Importancia de los distintos atributos. *Elaboración propia.*

Si observamos la comparativa entre la gráfica horaria de la demanda y el precio cuando normalizamos de la manera descrita en la Sección 4.2, vemos en la figura 4.14 que se ha perdido cualquier relación entre ambas variables. Por tanto, cabe plantearse qué ocurre cuando normalizamos ambas variables siguiendo el mismo método, análisis que se realiza en la siguiente subsección.

Tabla 4.8: Tabla del Error Absoluto Medio por mes y anual.

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
5.04	6.22	4.13	4.86	5.61	4.48	4.14	3.61	3.05	4.17	4.38	3.64	4.44

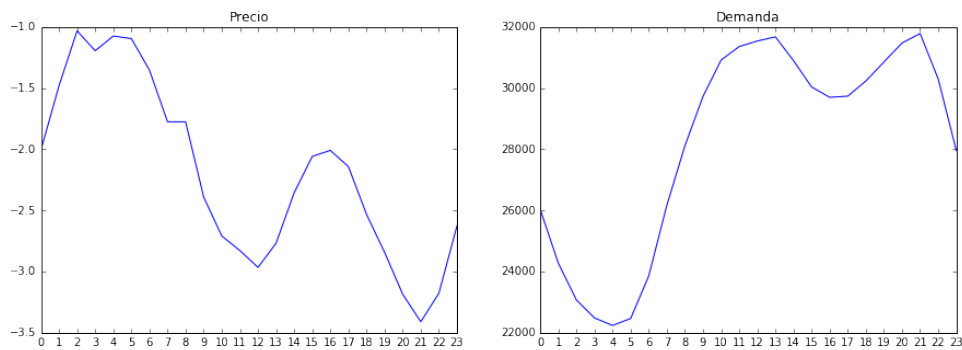


Figura 4.14: Comparativa entre la demanda y el precio normalizado. *Elaboración propia.*

RF con la demanda y el precio normalizados

Si volvemos a ejecutar la hiperparametrización normalizando también la demanda vemos que mejora algo los resultados con respecto al caso anterior no solo en total, también mes a mes, como podemos observar en la tabla 4.9.

Tabla 4.9: Tabla del Error Absoluto Medio por mes y anual para demanda y precio normalizados.

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
5.03	6.19	4.00	4.66	5.29	4.29	3.90	3.45	2.99	3.90	4.22	3.55	4.28

Además, si nos fijamos en la figura 4.15, donde se calculan las variables más importantes del mejor modelo, observamos que la demanda ahora se encuentra en sexta posición y la generación eólica y el precio 24 horas antes siguen en cabeza. De algún modo al normalizar ambas variables de la misma forma estamos preservando de mejor manera la estructura común que tenían y que se rompe en caso de no hacerlo.

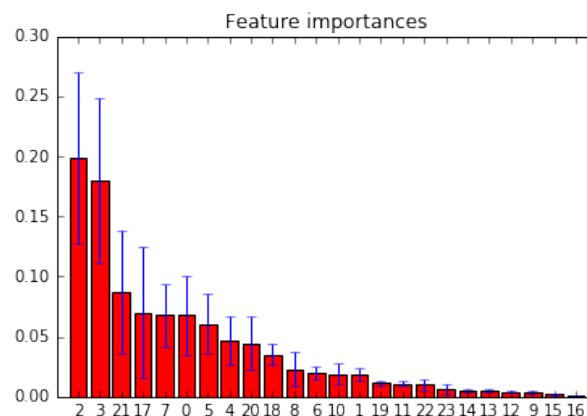


Figura 4.15: Importancia de los distintos atributos normalizando la demanda igual que los precios. *Elaboración propia.*

Si además de comprobar la importancia revisamos de nuevo la comparativa de las dos nuevas variables de demanda y precios, en la figura 4.16, observamos que aunque se preserve mucho más la relación horaria entre ambas, ésta es francamente mejorable si lo comparamos con la relación existente entre las dos variables sin normalizar como vemos en la figura 4.17. Debido a esto, se

realiza de nuevo otra hiperparametrización sin hacer uso de la normalización ni para la demanda ni para el precio.

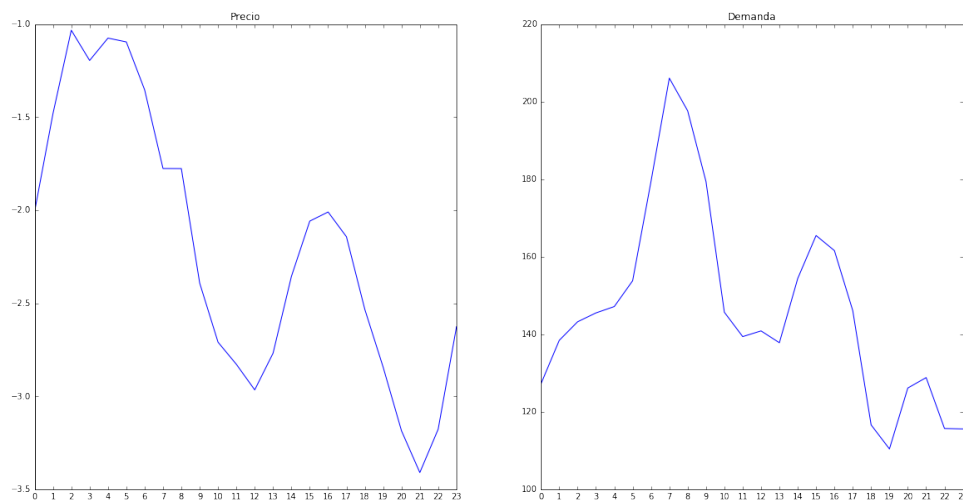


Figura 4.16: Comparativa entre la demanda y el precio, ambas variables normalizadas. *Elaboración propia.*

RF con datos sin normalizar

Al no normalizar, en la comparativa entre la evolución horaria de ambas variables de la figura 4.17 se observa la gran relación entre ambas comentada en el Capítulo 2. Con la hiperparametrización de *Random Forest* utilizando los datos sin normalizar observamos una mejoría notable. Por tanto, podemos concluir que la mejor opción es no normalizar al usar *Random Forest* para no perder ninguna relación subyacente entre los distintos atributos y el precio a pesar de que tengamos una mayor varianza en la variable objetivo.

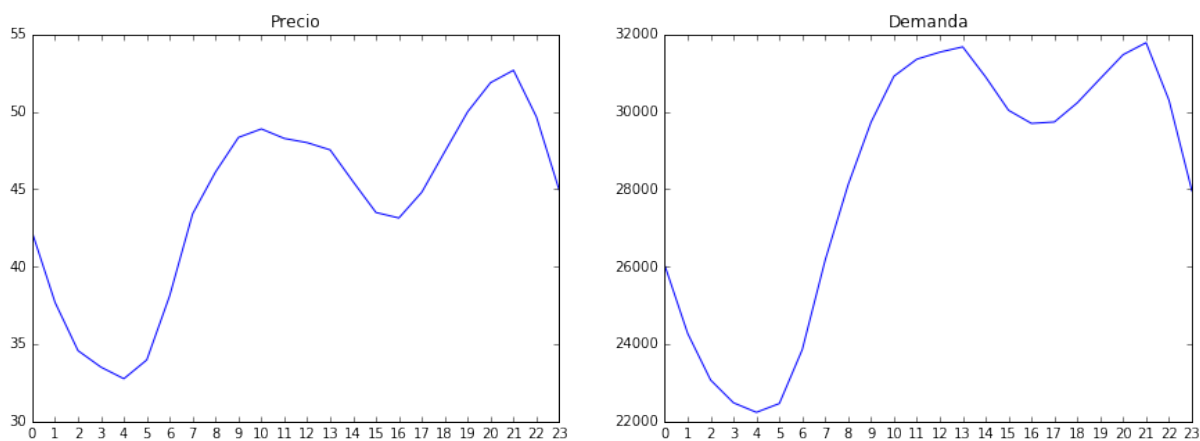


Figura 4.17: Comparativa entre la demanda y el precio, ambas variables sin normalizar. *Elaboración propia.*

En cuanto a la importancia de las variables, éstas han cambiado radicalmente. La demanda ahora se encuentra en primer lugar seguida del precio 24 horas antes y la generación de la eólica, y las demás tienen mucha menos relevancia, como podemos ver en la figura 4.18.

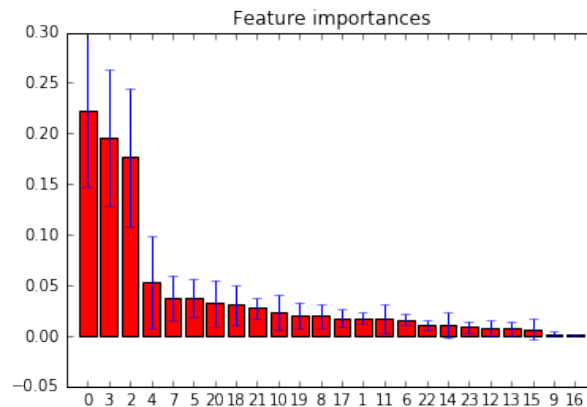


Figura 4.18: Importancia de las variables cuando no se normalizan. *Elaboración propia.*

Al añadir esta mejora podemos observar una mayor precisión en cada una de las horas de los meses, apreciable en la figura 4.19, donde observamos por un lado, que la línea verde se encuentra prácticamente por encima continuamente, indicando que el error es más alto prácticamente en todas las horas y en todos los meses mientras que la línea roja que representa al modelo que no normaliza las variables se encuentra normalmente por debajo de las otras dos, especialmente en abril, mayo, julio, agosto y diciembre.

En el resumen de todo el año que tenemos en la figura 4.20 se pueden observar tonos más suaves en el caso en el que se normaliza la demanda y aún más suaves en el mapa de calor correspondiente a los datos sin normalizar. Además, con estas figuras se nos permite entender que en general hay algunas horas y meses para los cuales es más difícil de predecir. Por ejemplo, en los meses de verano, sobre todo en agosto y septiembre, los errores son mucho más bajos para todas las horas del día que el resto del año. Este resultado coincide con el análisis de la Sección 2.3 donde se podía ver que en esos meses existe mucha menos variabilidad de precios tanto entre horas como entre distintos tipos de día.

Los histogramas de errores de la figura 4.22 y los gráficos de caja de la 4.21 nos permiten entender que en general en ninguno de los meses se está infraprediciendo ni sobre prediciendo. Es decir, todos los meses se encuentran más o menos centrados en el 0. Esto es así con las pequeñas excepciones de enero y febrero donde se sobrepredijo levemente y noviembre en el que ocurrió lo contrario. Pese a ello no se observa ningún sesgo que pueda ser corregida en un proceso posterior a la predicción.

En estos gráficos también se aprecia que la dispersión de los errores es aproximadamente constante, con una pequeña disminución en agosto y septiembre, provocada previsiblemente por el efecto del menor rango de valores que toma el precio en dichos meses.

Los errores del mejor modelo resultante, *Random Forest* sin normalización, se encuentran resumidos en la tabla 4.10. Son difíciles de comparar con los obtenidos por otros autores en parte porque un gran número de las publicaciones son lo suficientemente antiguas como para que el mercado haya cambiado profundamente hecho que ocurre en [20], [22] o [23]. De hecho la aparición de agentes participantes ha aumentado en gran cantidad en los últimos dos o tres años. Por otro lado, en numerosas publicaciones se toman medidas de ciertos meses, semanas o incluso días. Por ejemplo en [22] y en [23] mide el error usando semanas de distintos períodos. Por su parte, en [36] mide el error en ciertos días de junio. Incluso, otras publicaciones como [42], [20] realizan predicciones para otros mercados distintos al español y de difícil comparativa.

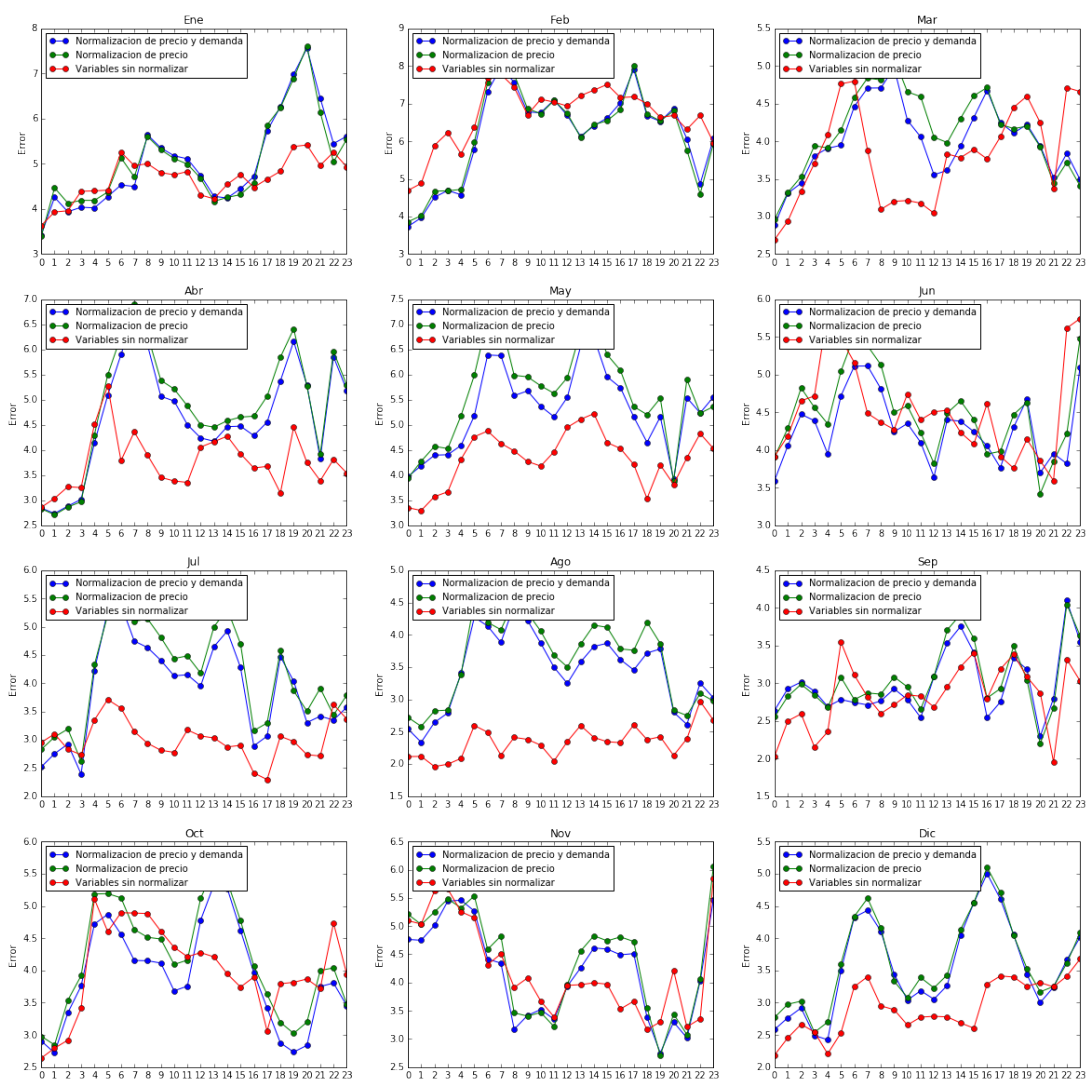


Figura 4.19: Errores por hora y mes. *Elaboración propia.*

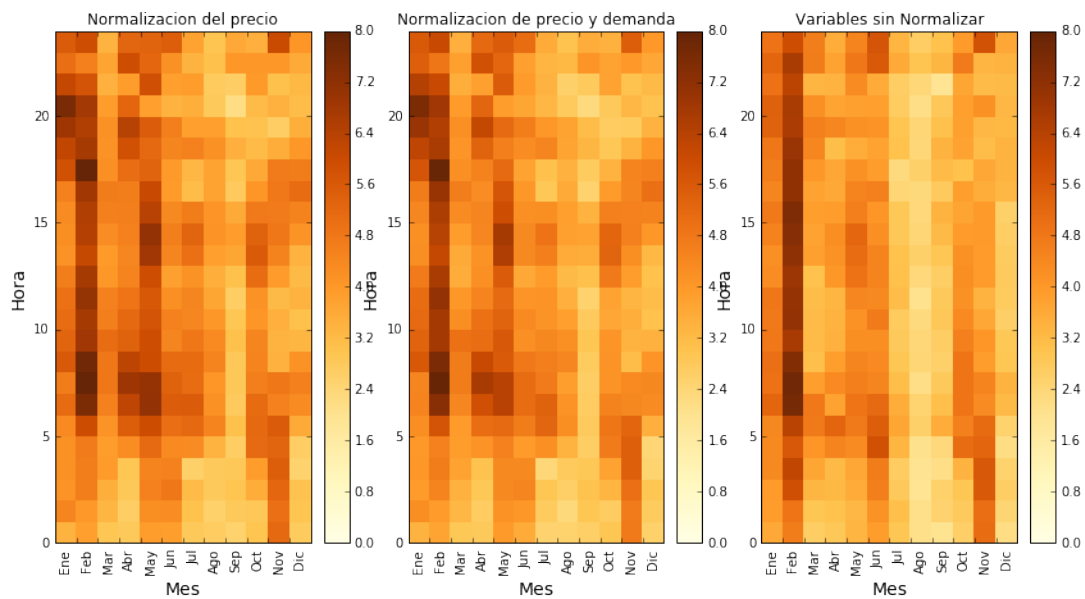


Figura 4.20: Resumen de errores por hora y mes para el año completo. *Elaboración propia.*

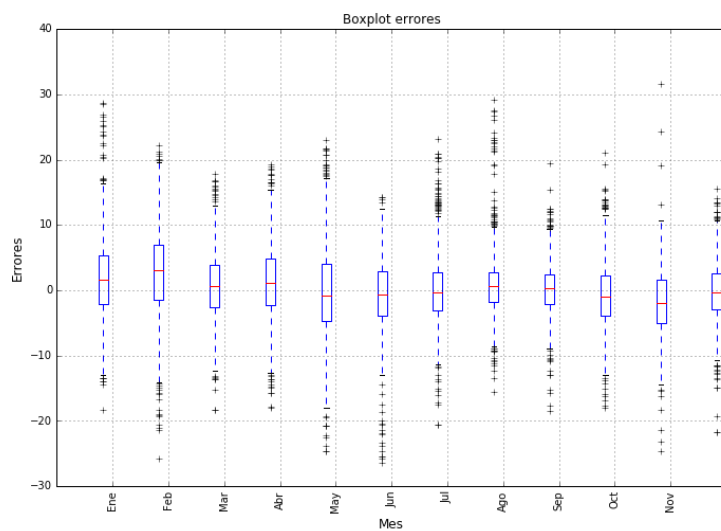
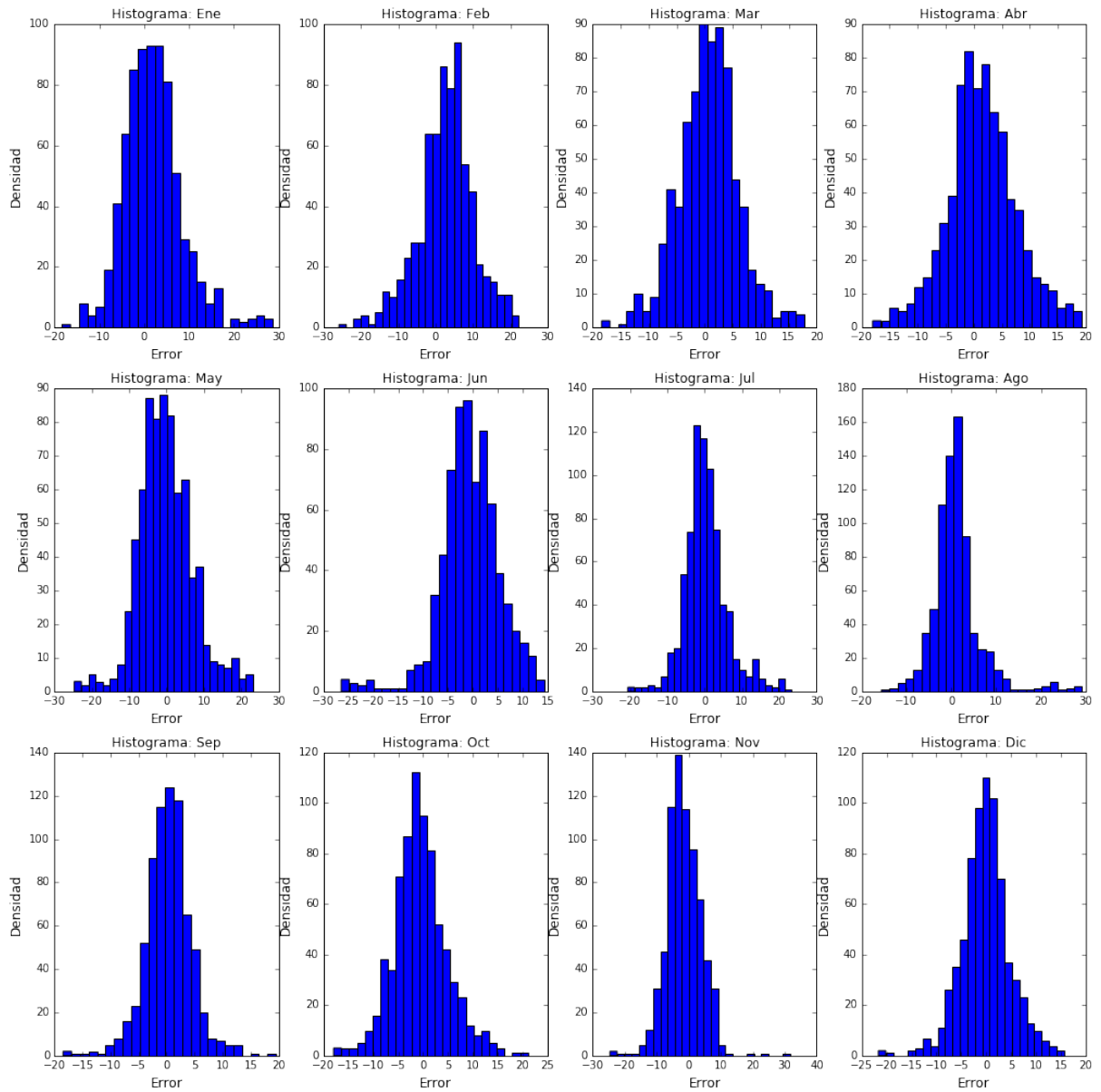


Figura 4.21: BoxPlot de errores con signo. *Elaboración propia.*

Figura 4.22: Histogramas de errores con signo. *Elaboración propia.*

Por toda esta problemática se convierte en muy difícil la comparativa eficaz con otras publicaciones anteriores y es por ello que solo nos queda comparar los modelos entre sí en un *benchmark* interno que de alguna manera servirá de modelo de referencia para los años más recientes.

Tabla 4.10: Tabla del Error Absoluto Medio por mes y anual para el mejor modelo: *Random Forest* con datos sin normalizar

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
4.67	6.67	3.80	3.75	4.32	4.53	3.00	2.34	2.83	4.01	4.24	2.93	3.92

4.6.2. Resultados de Test de Random Forest sin normalización

Como *test* se ha reservado el período disponible de 2017, es decir, de enero a julio, ambos inclusive. En él, se ha probado el uso del mejor modelo, *Random Forest* sin normalizar, de manera que se entrena cada día y se predice para el día siguiente. Siguiendo esta metodología, el resultado es algo peor que para el período de evaluación como es habitual en este tipo de problemas. Esto es debido a que de alguna manera al realizar la hiperparametrización se está ajustando el modelo para aprenderse el conjunto de validación.

Más en detalle el error de test es de 4,50€ y desglosado por meses lo podemos ver en la tabla 4.11 donde observamos que el error se degrada sobre todo a partir del mes de abril. A pesar de esto, en la operativa real de un producto como este, existen dos elementos de mantenimiento del modelo que harían mejorar sus predicciones. En primer lugar no se tendría un único modelo, sino que habría varios que se encuentren en seguimiento y que de forma periódica podrían cambiar porque haya algún modelo que esté ofreciendo mejores resultados que el que se encuentre en producción. En segundo lugar, que a partir del mes de abril el modelo empeore puede deberse a una necesidad de rehiperparametrización. En un servicio como este sería vital definir una hiperparametrización periódica que solviera dicho problema.

Tabla 4.11: Tabla del Error Absoluto Medio por mes y medio para el mejor modelo en test: *Random Forest* con datos sin normalizar

Ene	Feb	Mar	Abr	May	Jun	Jul	Medio
3.30	2.92	4.60	6.28	4.60	8.01	3.25	4.50

4.7. Modelo de explotación

Por su naturaleza y por cómo está formado el mercado eléctrico en España, un modelo de predicción de precios es muy interesante para empresas generadoras de energía eléctrica, comercializadoras y *traders* de mercado. Las empresas generadoras, pueden decidir parar su producción en función del precio al que se vaya a pagar la energía en el mercado, las empresas comercializadoras, sobre todo aquellas que trabajan con grandes industrias, modifican su estrategia de compra en función del precio y, por su puesto, sería de vital importancia para los *traders* que se pueden situar a un lado o a otro de la oferta y la demanda.

Dada la gran cantidad de empresas que pueden estar interesadas en un producto que oferte este tipo de predicciones y a su heterogeneidad, lo más adecuado es que el modelo de explotación del mismo se realice mediante un *Software as a Service* (SaaS) que es un modelo mediante el

cual los datos, el código y los recursos necesarios para realizar las predicciones se encuentran del lado del proveedor y se podrá acceder vía una página web a toda la información relevante. Este tipo de modelo de explotación permite al cliente no preocuparse por posibles necesidades de actualizaciones del modelo, ni por los recursos materiales como servidores o licencias, ni por los técnicos necesarios para el correcto funcionamiento del modulo de predicción. Por su parte, el proveedor mantiene la propiedad intelectual de los desarrollos pudiendo ofertar el producto a distintos potenciales clientes.

El servicio de predicción seguiría el diseño resumido en la figura 4.23. En este diseño, las variables de entrada se obtienen en primer lugar a través del portal de datos abiertos de REE, E-SIOS, donde se obtienen datos de demanda, de producción eólica y solar, de precios de la electricidad tanto en España como de Francia. La obtención de los datos se realiza a partir de la API que ofrece REE de manera gratuita. En segundo lugar, se necesitan datos de las festividades que se deben obtener del BOE y pueden variar de un año a otro y se publican en una tabla en formato PDF que es difícil de procesar. Es por ello que este proceso deberá ser mayoritariamente manual pero solo se necesitará realizar anualmente. Por último, se necesitan variables meteorológicas tanto para las predicciones de demanda como de producción eólica y solar. Estas meteorologías pueden ser obtenidas mediante entidades o compañías de terceros como bien pueden ser The European Centre for Medium-Range Weather Forecasts (ECMWF) [3], The Weather Company (TWC) [12] de IBM o Global Forecast System (GFS) [6], para ello, se debe elegir y contratar una rejilla que cubra al menos el territorio español.

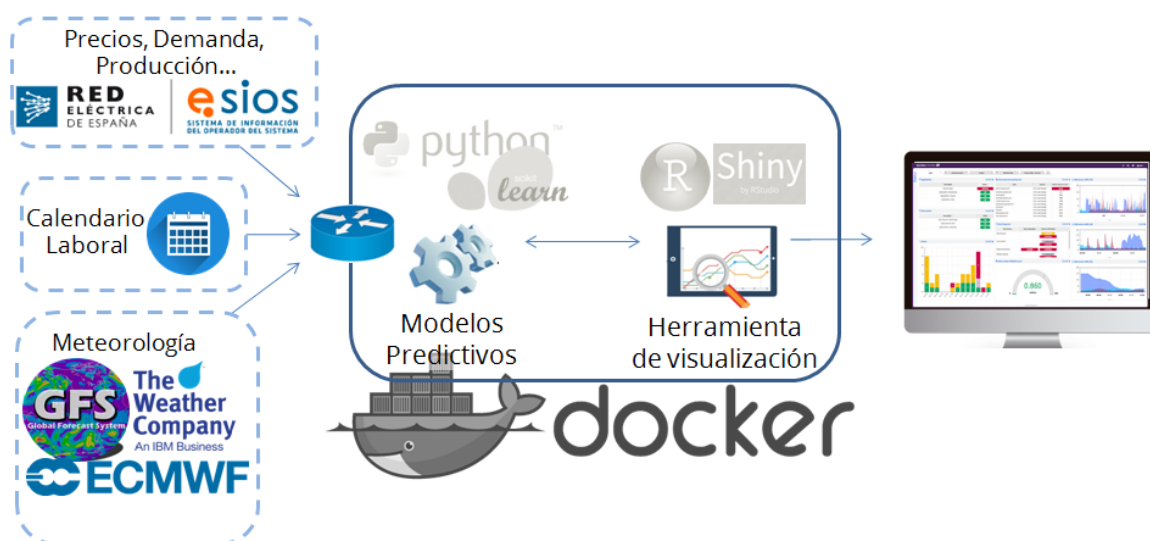


Figura 4.23: Diseño del desarrollo de la aplicación. *Elaboración propia.*

El diseño contempla que el motor de predicción estará implementado en Python, en concreto haciendo uso de la potente librería Scikit-Learn [9] que contiene muchas facilidades para la creación de modelos predictivos. Sin embargo, gracias a la potencia para la visualización que posee el lenguaje de programación R, se contempla que la herramienta de visualización que permitirá al usuario, además de consultar, mediante un *dashboard*, el precio horario para el día siguiente, consultar estadísticas de la serie temporal de precios y del propio modelo de predicción y que podrá ser personalizable si el cliente así lo necesita esté desarrollada en R. La librería Shiny [10] y ShinyDashboard [11] de este lenguaje son las más apropiadas para conseguir una visualización adecuada y atractiva para los usuarios.

La decisión de dividir el servicio en dos entidades, una desarrollada en Python y la otra en R se debe a que, como ya hemos apuntado, aunque ambos lenguajes son los mejores para desarrollar algoritmos predictivos y para proyectos de Ciencia de Datos, la librería de Scikit-Learn es muy completa y ofrece un marco común a muy diversos modelos y, sobre todo, ofrece algunos elementos diferenciadores para desarrollar un código de calidad que será mantenido y que se va a encontrar en producción. Estas opciones son la facilidad para el *testing*, uso de excepciones y de clases, todo esto bajo un lenguaje sencillo de entender porque pretende ser lo más cercano posible al pseudocódigo. Además, es un lenguaje mucho más eficiente y que puede ser compilado, lo cual permite incluso una implantación en cliente sin perder el control del código de forma que se simplifica el mantenimiento de la propiedad intelectual.

Por otro lado, para el desarrollo de la herramienta de visualización, existirían otras alternativas como usar librerías como D3 o incluso HTML y Javascript de forma directa pero esta decisión aunque ciertamente podría ofrecer más potencia en cuanto a las posibilidades de visualización, también implicaría un mayor coste y tiempo de implementación y de mantenimiento. La opción aquí planteada haciendo uso de Shiny y ShinyDashboard permite que ante cambios requeridos por los potenciales clientes, la respuesta sea mucho más rápida, generando una mayor confianza y satisfacción en ellos.

Todo el código se encontrará dentro de un contenedor Docker [2] que permita el despliegue de la aplicación en producción lo más sencillo y rápido posible. Docker permite crear contenedores que siguen la misma filosofía que los contenedores físicos que sirven para transportar mercancías. Estos contenedores están estandarizados y todos tienen las mismas medidas independientemente del contenido, facilitando así el transporte de distintas mercancías ya sea por barco, tren, etc. En el caso de Docker permite que el servidor en el que sea instalada la aplicación no necesite conocer los paquetes, librerías y otros elementos que deben ser instalados para el correcto funcionamiento sino que solamente hay que tener instalado Docker y desplegar el contenedor que ya contiene todo el software y librerías necesarias. El uso de Docker permitirá, por tanto, aislarse del resto del servidor de forma similar a como lo haría una máquina virtual pero con muchos mejores resultados en eficiencia.

5

Conclusiones y trabajo futuro

“Data is just like crude. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analysed for it to have value.”

Michael Palmer

5.1. Conclusiones

En este trabajo se han implementado distintos modelos con el objetivo final de resolver el problema de la predicción de precios del mercado eléctrico. Para ello, se han seguido varias etapas. En primer lugar ha sido necesario revisar de forma exhaustiva el funcionamiento de el sistema eléctrico en su conjunto y, en concreto, su mercado para comprender cómo se determinan los precios en él, quiénes son los posibles agentes interesados en este tipo de modelos predictivos y poder definir de forma correcta un servicio que resuelva la problemática.

En segundo lugar, se han estudiado los distintos modelos y variables, presentes en la bibliografía que han sido utilizados para la predicción de precios. Cabe destacar la dificultad en la comparativa con las publicaciones citadas en este trabajo debido a la diferencia temporal y geográfica de las publicaciones con respecto a mi trabajo.

Para la elección del modelo se ha llevado a cabo un *benchmark* entre distintos tipos de modelos y parámetros entre los que se encuentran la Regresión Lineal Ridge, K-Vecinos Próximos, Perceptrón Multicapa, Máquina de Vectores Soporte y *Random Forest*. De entre todas las combinaciones de modelos y parámetros el más preciso ha sido *Random Forest* y para él se ha realizado un estudio más detallado, incluyendo un análisis de las variables más relevantes, una comparativa de distintos tipos de normalización, una exploración de los errores por meses, horas, etc.

Dentro de los apartados más claves en esta sección es tanto la importancia de una correcta hiperparametrización como la normalización. Hemos visto que si normalizamos la serie temporal de precios para reducir su varianza puede provocar una pérdida de información sobre los patrones subyacente en los datos que es de gran utilidad para la predicción. En concreto, se ha observado como la normalización difuminaba la gran relación que tiene el recio con la demanda, que se soluciona parcialmente aplicando el mismo tratamiento de normalización a ambas variables. A pesar de esta mejora, los resultados aportan un resultado desfavorable a la normalización ya que sin ella se consiguen los mejores resultados como se puede observar en la tabla 5.1, donde:

- RF - 1: modelo de *Random Forest* con solamente el precio normalizado.
- RF - 2: modelo de *Random Forest* con el precio y la demanda normalizados.
- RF - 3: modelo de *Random Forest* sin normalizar ninguno de los datos.

El mejor modelo predictivo conseguido, ha conseguido obtener un MAE en validación 3,92 y de 4,50 en *test*, resultado que ya es de utilidad para los todos los agentes que participan en el mercado.

Tabla 5.1: Resumen de resultados con distintas aproximaciones.

Modelo	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Total
RF - 1	5.04	6.22	4.13	4.86	5.61	4.48	4.14	3.61	3.05	4.17	4.38	3.64	4.44
RF - 2	5.03	6.19	4.00	4.66	5.29	4.29	3.90	3.45	2.99	3.90	4.22	3.55	4.28
RF - 3	4.67	6.67	3.80	3.75	4.32	4.53	3.00	2.34	2.83	4.01	4.24	2.93	3.92

Por último, se ha realizado una propuesta de explotación como *SaaS* que ofrece beneficios tanto al proveedor como al cliente y se ha definido el diseño técnico de la aplicación para su puesta en producción. En esta definición se ha decidido el uso de Python y la librería Scikit-Learn como base del motor de predicción; R con sus librerías Shiny y ShinyDashboard para la implementación de una interfaz web que muestre los resultados de forma amigable y sencilla y, Docker para encapsular la solución y facilitar su despliegue en cualquier servidor y su integración con otros módulos en caso de ser necesario, convirtiéndolo, de algún modo, en una pieza independiente del contexto en el que se encuentra.

Este servicio de predicción es un sistema muy útil y de gran interés para los distintos agentes del mercado pudiendo servir como herramienta de ayuda a la toma de decisiones en el complejo mercado eléctrico.

Como resultado de este trabajo, surgen dos iniciativas en desarrollo, un nuevo servicio de predicción de precios derivado de este trabajo que se incluirá dentro de la oferta tecnológica del IIC y una publicación en una revista internacional para marzo de 2018 [40].

5.2. Trabajo futuro

La evolución del precio se debe en gran medida a los factores que han sido analizados en este trabajo pero además de estos existen otros más complejos que no han sido tenidos en cuenta en este trabajo y que serían posibles líneas futuras de investigación.

Entre estas variables a investigar, destaca la influencia de factores económicos, entre los que se pueden incluir el precio de las materias primas más usadas para la obtención de energía (carbón, petróleo y gas), el Producto Interior Bruto, etc.

Otro aspecto muy importante a considerar son las interconexiones con los países limítrofes, en el caso de España. La conexión más importante es la francesa y, por ello, se incluyó el precio en Francia pero además de este precio, es muy importante la capacidad técnica de exportación e importación así como las cantidades importadas/exportadas.

La laboralidad es un aspecto clave en la predicción de precios debido al efecto que tienen los festivos, sábados y domingos en la demanda de energía eléctrica. Por tanto, este es un aspecto clave que se debe seguir investigando. En este trabajo se ha intentado solucionar mediante una normalización que se ha descubierto como poco efectiva y posteriormente como variables que usa el propio *Random Forest* aunque les dé poca importancia. Por tanto, otras líneas de investigación en este sentido son, por un lado, realizar un postproceso de los días clave (festivos, época de Navidad y de verano, puentes, etc.) y, por otro, realizar modelos para esos días por separado, con la problemática de la escasez de datos.

Por último, el aspecto más importante y a la vez el más difícil de incluir en el modelo es la estrategia de los agentes en el mercado. Pese a su dificultad hay datos que son publicados por el OMIE que se podrían utilizar para solventar esta problemática.

Acrónimos

- **REE** Red Eléctrica Española
- **EDP** Energía de Portugal
- **OMIE** Operador del Mercado Ibérico de Energía
- **OMIP** Operador del Mercado Ibérico de Energía - Polo Portugués
- **CF** Coeficiente de Festividad
- **AR** Autorregresive Models
- **MA** Moving Average Models
- **ACF** Autocorrelation Function
- **SVM** Support Vector Machine
- **SVR** Support Vector Machine for Regression
- **MLP** Multiplayer Perceptron
- **K-NN** K-Nearest Neighbours
- **RF** Random Forest
- **MAE** Mean Absolute Error
- **MSE** Mean Squared Error
- **MSD** Mean Signed Deviation
- **SaaS** Software as a Service
- **ECMWF** The European Centre for Medium-Range Weather Forecasts
- **TWC** The Weather Company
- **GFS** Global Forecast System

Bibliografía

- [1] 1.6. nearest neighbors — scikit-learn 0.19.0 documentation. <http://scikit-learn.org/stable/modules/neighbors.html>. (Accessed on 09/10/2017).
- [2] Docker - build, ship, and run any app, anywhere. <https://www.docker.com/>. (Accessed on 09/10/2017).
- [3] Ecmwf. <https://www.ecmwf.int/>. (Accessed on 09/10/2017).
- [4] El precio de la luz se dispara este miércoles en plena ola de frío | economía home | el mundo. <http://www.elmundo.es/economia/2017/01/18/587e7f3c468aeb09068b4641.html>. (Accessed on 09/09/2017).
- [5] Esios electricidad · datos · transparencia. <https://www.esios.ree.es/es>. (Accessed on 09/09/2017).
- [6] Global forecast system (gfs) | national centers for environmental information (ncei) formerly known as national climatic data center (ncdc). <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>. (Accessed on 09/10/2017).
- [7] Las cementeras advierten de posibles paros de fábricas por el precio de la luz | economía | el país. https://economia.elpais.com/economia/2017/01/25/actualidad/1485345243_343311.html. (Accessed on 09/09/2017).
- [8] Regression with residuals. <http://pgfplots.net/tikz/examples/regression-residuals/>. (Accessed on 09/10/2017).
- [9] Scikit-learn: machine learning in python. <http://scikit-learn.org/stable/>. (Accessed on 09/10/2017).
- [10] Shiny. <https://shiny.rstudio.com/>. (Accessed on 09/10/2017).
- [11] Shiny dashboard. <https://rstudio.github.io/shinydashboard/>. (Accessed on 09/10/2017).
- [12] The weather company | an ibm business. <http://www.theweathercompany.com/>. (Accessed on 09/10/2017).
- [13] Sanjeev Kumar Aggarwal, Lalit Mohan Saini, and Ashwani Kumar. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems*, 31(1):13–22, 2009.
- [14] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, September 1975.

- [15] Christopher M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- [16] George E. P. Box and Gwilym M. Jenkins. *Time series analysis: forecasting and control*. Holden-Day, 1976.
- [17] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- [18] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [19] Derek W. Bunn. Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE*, 88(2):163–169, 2000.
- [20] Jinxing Che and Jianzhou Wang. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Conversion and Management*, 51(10):1911–1917, 2010.
- [21] Xia Chen, Zhao Yang Dong, Ke Meng, Yan Xu, Kit Po Wong, and HW Ngan. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Transactions on Power Systems*, 27(4):2055–2062, 2012.
- [22] Antonio J. Conejo, Miguel A. Plazas, Rosa Espinola, and Ana B. Molina. Day-ahead electricity price forecasting using the wavelet transform and arima models. *IEEE transactions on power systems*, 20(2):1035–1042, 2005.
- [23] Javier Contreras, Rosario Espinola, Francisco J. Nogales, and Antonio J. Conejo. Arima models to predict next-day electricity prices. *IEEE transactions on power systems*, 18(3):1014–1020, 2003.
- [24] Alberto Cruz, Antonio Muñoz, Juan Luis Zamora, and Rosa Espínola. The effect of wind generation and weekday on spanish electricity spot price forecasting. *Electric Power Systems Research*, 81(10):1924–1935, 2011.
- [25] Jonathan D. Cryer and Natalie Kellet. *Time series analysis*, volume 101. Springer, 1986.
- [26] Jesús Crespo Cuaresma, Jaroslava Hlouskova, Stephan Kossmeier, and Michael Obersteiner. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, 77(1):87–106, 2004.
- [27] Raquel Gareta, Luis M. Romeo, and Antonia Gil. Forecasting of electricity prices with neural networks. *Energy Conversion and Management*, 47(13):1770–1778, 2006.
- [28] Manuel Gurrea. Análisis de componentes principales. *Proyecto e-Math Financiado por la Secretaría de Estado de Educación y Universidades (MECD)*, 2000.
- [29] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [30] Z. Hu, Y. Yu, Z. Wang, W. Sun, D. Gan, and Z. Han. Price forecasting using an integrated approach. In *Electric Utility Deregulation, Restructuring and Power Technologies, 2004.(DRPT 2004). Proceedings of the 2004 IEEE International Conference on*, volume 1, pages 28–31. IEEE, 2004.
- [31] G Jay Kerns. *Introduction to probability and statistics using R*. Lulu. com, 2010.

- [32] Chang-il Kim, In-Keun Yu, and YH Song. Prediction of system marginal price of electricity using wavelet transform analysis. *Energy Conversion and Management*, 43(14):1839–1851, 2002.
- [33] Hyafil Laurent and Rivest Ronald. Constructing optimal binary decision tress is np-complete. *Information Processing Letters*, 1976.
- [34] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.
- [35] Alicia Troncoso Lora, Jose Riquelme Santos, Jesús Riquelme Santos, José Luís Martínez Ramos, and Antonio Gómez Expósito. Electricity market price forecasting: Neural networks versus weighted-distance k nearest neighbours. In *International Conference on Database and Expert Systems Applications*, pages 321–330. Springer, 2002.
- [36] Jie Mei, Dawei He, Ronald Harley, Thomas Habetler, and Guannan Qu. A random forest method for real-time price forecasting in new york electricity market. In *PES General Meeting/ Conference & Exposition, 2014 IEEE*, pages 1–5. IEEE, 2014.
- [37] Hiroyuki Mori and Akira Awata. Data mining of electricity price forecasting with regression tree and normalized radial basis function network. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pages 3743–3748. IEEE, 2007.
- [38] Stephen M. Omohundro. Five balltree construction algorithms. Technical report, 1989.
- [39] Lior Rokach and Oded Maimon. *Data mining with decision trees: theory and applications*. World scientific, 2014.
- [40] Álvaro Romero-Miralles, Jose R. Dorronsoro-Ibero, and Julia Díaz-García. Day-ahead price forecasting for spanish electricity market. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2018.
- [41] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [42] Damien C. Sansom, Tom Downs, Tapan K. Saha, et al. Evaluation of support vector machine based forecasting tool in electricity price forecasting for australian national electricity market participants. *Journal of Electrical & Electronics Engineering, Australia*, 22(3):227, 2003.
- [43] Robert E. Schapire, Yoav Freund, Peter Bartlett, Wee Sun Lee, et al. Boosting the margin: A new explanation for the effectiveness of voting methods. *The annals of statistics*, 26(5):1651–1686, 1998.
- [44] Robert H. Shumway and David S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.
- [45] BR Szkuta, L. Augusto Sanabria, and Tharam S. Dillon. Electricity price short-term forecasting using artificial neural networks. *IEEE transactions on power systems*, 14(3):851–857, 1999.
- [46] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.
- [47] Paul J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, 1974.

- [48] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International journal of forecasting*, 30(4):1030–1081, 2014.
- [49] Rafał Weron and Adam Misiorek. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International journal of forecasting*, 24(4):744–763, 2008.
- [50] Rafal Weron, Adam Misiorek, et al. Forecasting spot electricity prices with time series models. In *Proceedings of the European Electricity Market EEM-05 Conference*, pages 133–141, 2005.
- [51] G. Udny Yule. On the time-correlation problem, with especial reference to the variate-difference correlation method. *Journal of the Royal Statistical Society*, 84(4):497–537, 1921.
- [52] G. Udny Yule. On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226:267–298, 1927.
- [53] Jun Hua Zhao, Zhao Yang Dong, Zhao Xu, and Kit Po Wong. A statistical approach for interval forecasting of the electricity price. *IEEE Transactions on Power Systems*, 23(2):267–276, 2008.

Apéndices



Tablas de resultados

En este anexo se incluyen los principales resultados de todos los modelos probados. Más concretamente se especifican los parámetros que mejor han funcionado en la predicción del precio detallada en el Capítulo 4

alpha	trainDays	norm	mae	mse
1.0100	90	maxAbs	5.7531	54.9769
1	90	maxAbs	5.7533	54.9832
0.9900	90	maxAbs	5.7535	54.9896
0.9800	90	maxAbs	5.7538	54.9959
0.9700	90	maxAbs	5.7540	55.0024
0.9600	90	maxAbs	5.7542	55.0088
0.9500	90	maxAbs	5.7544	55.0154
1.0100	90	minMax	5.7676	55.2514
1	90	minMax	5.7677	55.2557
0.9900	90	minMax	5.7678	55.2601
0.9800	90	minMax	5.7679	55.2645
0.9700	90	minMax	5.7680	55.2689
0.9600	90	minMax	5.7681	55.2734
0.9500	90	minMax	5.7682	55.2779
1.0100	90	robust	5.7798	55.9336
1	90	robust	5.7798	55.9348
0.9900	90	robust	5.7799	55.9361
0.9800	90	robust	5.7799	55.9373
0.9700	90	robust	5.7800	55.9386
0.9600	90	robust	5.7801	55.9398
0.9500	90	robust	5.7801	55.9411
1.0100	90	standard	5.7848	56.0502
1	90	standard	5.7848	56.0513

alpha	trainDays	norm	mae	mse
0.9900	90	standard	5.7849	56.0525
0.9800	90	standard	5.7849	56.0536
0.9700	90	standard	5.7849	56.0548
0.9600	90	standard	5.7850	56.0560
0.9500	90	standard	5.7850	56.0572
1.0100	150	maxAbs	5.9094	56.8871
1	150	maxAbs	5.9095	56.8911
0.9900	150	maxAbs	5.9097	56.8951
0.9800	150	maxAbs	5.9098	56.8991
0.9700	150	maxAbs	5.9100	56.9032
0.9600	150	maxAbs	5.9101	56.9072
0.9500	150	maxAbs	5.9102	56.9113
1.0100	150	minMax	5.9199	57.0681
1	150	minMax	5.9200	57.0705
0.9900	150	minMax	5.9200	57.0730
0.9800	150	minMax	5.9201	57.0754
0.9700	150	minMax	5.9201	57.0779
0.9600	150	minMax	5.9202	57.0803
0.9500	150	minMax	5.9202	57.0828
0.9500	150	standard	5.9234	57.3790
0.9600	150	standard	5.9234	57.3800
0.9700	150	standard	5.9235	57.3810
0.9800	150	standard	5.9235	57.3820
0.9900	150	standard	5.9236	57.3830
1	150	standard	5.9236	57.3839
1.0100	150	standard	5.9236	57.3848
0.9500	150	robust	5.9238	57.4070
0.9600	150	robust	5.9238	57.4077
0.9700	150	robust	5.9239	57.4084
0.9800	150	robust	5.9239	57.4091
0.9900	150	robust	5.9239	57.4098
1	150	robust	5.9239	57.4104
1.0100	150	robust	5.9239	57.4110
1.0100	30	maxAbs	6.0655	67.3848
1	30	maxAbs	6.0674	67.4502
0.9900	30	maxAbs	6.0692	67.5163
0.9800	30	maxAbs	6.0711	67.5832
0.9700	30	maxAbs	6.0730	67.6509
0.9600	30	maxAbs	6.0749	67.7193
0.9500	30	maxAbs	6.0768	67.7886
0.9500	210	standard	6.1007	60.0057
0.9600	210	standard	6.1009	60.0105
0.9700	210	standard	6.1012	60.0152
0.9800	210	standard	6.1014	60.0198
0.9900	210	standard	6.1016	60.0244
1	210	standard	6.1018	60.0288

alpha	trainDays	norm	mae	mse
1.0100	210	standard	6.1020	60.0332
1.0100	210	maxAbs	6.1040	59.9379
1	210	maxAbs	6.1041	59.9414
0.9900	210	maxAbs	6.1043	59.9450
0.9800	210	maxAbs	6.1044	59.9485
0.9700	210	maxAbs	6.1046	59.9521
0.9600	210	maxAbs	6.1047	59.9557
0.9500	210	maxAbs	6.1048	59.9592
0.9500	210	robust	6.1091	60.2307
0.9600	210	robust	6.1093	60.2335
0.9700	210	robust	6.1094	60.2363
0.9800	210	robust	6.1095	60.2390
0.9900	210	robust	6.1097	60.2417
1	210	robust	6.1098	60.2443
1.0100	210	robust	6.1099	60.2468
1.0100	210	minMax	6.1133	60.1124
1	210	minMax	6.1134	60.1142
0.9900	210	minMax	6.1134	60.1161
0.9800	210	minMax	6.1134	60.1180
0.9700	210	minMax	6.1135	60.1199
0.9600	210	minMax	6.1135	60.1218
0.9500	210	minMax	6.1136	60.1237
0.9500	270	standard	6.2747	63.0568
0.9600	270	standard	6.2751	63.0659
0.9700	270	standard	6.2755	63.0749
0.9800	270	standard	6.2759	63.0836
0.9900	270	standard	6.2763	63.0923
1	270	standard	6.2768	63.1008
1.0100	270	standard	6.2772	63.1091
0.9500	270	robust	6.2945	63.4923
0.9600	270	robust	6.2948	63.4981

Tabla A.1: Resultados hiperparametrización de Regresión Lineal Ridge.

n_neighbors	trainDays	norm	mae	mse
20	270	robust	5.3716	52.5299
30	270	robust	5.4064	52.7941
10	270	robust	5.4279	53.6369
40	270	robust	5.4464	53.2482
50	270	robust	5.5009	54.1905
10	210	robust	5.5293	55.4839
20	210	robust	5.5407	55.5156
60	270	robust	5.5530	55.1356
10	270	standard	5.5600	56.8833

n_neighbors	trainDays	norm	mae	mse
20	270	standard	5.5627	56.0879
30	210	robust	5.5681	55.6080
10	150	robust	5.5926	57.5920
70	270	robust	5.5953	55.8952
40	210	robust	5.6059	56.4000
80	270	robust	5.6381	56.7582
20	150	robust	5.6455	57.6275
30	270	standard	5.6516	57.2039
10	210	standard	5.6624	59.0607
50	210	robust	5.6658	57.4770
10	90	robust	5.6749	59.7133
90	270	robust	5.6777	57.4616
20	210	standard	5.6825	58.3126
20	90	robust	5.6968	59.5486
30	150	robust	5.7011	58.5545
60	210	robust	5.7126	58.3734
40	270	standard	5.7246	58.2453
30	210	standard	5.7423	58.6062
30	90	robust	5.7511	60.1589
40	150	robust	5.7543	59.7560
70	210	robust	5.7574	59.2025
50	270	standard	5.7843	59.4373
40	210	standard	5.7978	59.6366
40	90	robust	5.8058	61.2089
80	210	robust	5.8073	60.1549
50	150	robust	5.8076	60.7401
10	150	standard	5.8383	65.0932
60	270	standard	5.8385	60.5880
60	150	robust	5.8565	61.5728
50	210	standard	5.8574	60.8621
90	210	robust	5.8620	61.2202
50	90	robust	5.8632	62.1346
20	150	standard	5.8734	65.1544
70	270	standard	5.8932	61.7078
30	150	standard	5.8944	65.1329
60	210	standard	5.9046	61.9872
70	150	robust	5.9190	62.5857
60	90	robust	5.9318	63.2809
80	270	standard	5.9396	62.7156
40	150	standard	5.9406	66.1391
70	210	standard	5.9593	63.1757
40	90	standard	5.9701	66.2701
90	270	standard	5.9802	63.5974
30	90	standard	5.9812	66.5913
80	150	robust	5.9813	63.6630
50	150	standard	5.9894	67.2382

n_neighbors	trainDays	norm	mae	mse
10	90	standard	5.9901	68.1200
70	90	robust	5.9914	64.3552
10	270	minMax	5.9947	65.1836
50	90	standard	6.0041	66.6526
80	210	standard	6.0043	64.1450
20	90	standard	6.0050	68.1143
10	30	robust	6.0167	66.6007
20	30	robust	6.0288	66.3652
90	210	standard	6.0440	65.0365
90	150	robust	6.0460	64.8487
60	150	standard	6.0464	68.1672
80	90	robust	6.0531	65.6225
60	90	standard	6.0688	67.4785
30	30	robust	6.0871	66.7650
10	270	maxAbs	6.0929	66.6083
70	150	standard	6.0937	68.7606
90	90	robust	6.1025	66.7303
20	270	minMax	6.1041	65.7588
70	90	standard	6.1185	68.1877
20	270	maxAbs	6.1347	66.0760
80	150	standard	6.1364	69.2867
40	30	robust	6.1584	67.9449
80	90	standard	6.1636	69.0358
90	150	standard	6.1761	69.7320
10	210	minMax	6.1849	70.3828
30	270	minMax	6.1878	66.9557
30	270	maxAbs	6.2039	67.1309
90	90	standard	6.2102	70.0226
50	30	robust	6.2346	69.3443
10	210	maxAbs	6.2485	70.8067
20	210	minMax	6.2615	70.9086
40	270	minMax	6.2901	68.9609
20	210	maxAbs	6.2933	70.8215
40	270	maxAbs	6.3073	68.8853
60	30	robust	6.3112	71.1310
30	210	minMax	6.3524	72.2491
10	30	standard	6.3616	76.9546
30	210	maxAbs	6.3710	72.1587
10	150	minMax	6.3778	77.4702
10	150	maxAbs	6.3801	76.6099
70	30	robust	6.3840	73.1434
50	270	minMax	6.4020	70.8990
50	270	maxAbs	6.4275	71.1019

Tabla A.2: Resultados hiperparametrización de K-Vecinos Próximos.

n_estimators	min_samp_leaf	max_feat	trainDays	mae	mse
710	2	0.4	270	4.5148	37.3080
610	2	0.4	270	4.5154	37.2816
910	2	0.4	270	4.5157	37.3582
810	2	0.4	270	4.5169	37.3762
310	2	0.4	270	4.5170	37.2798
510	2	0.4	270	4.5180	37.3188
410	2	0.4	270	4.5190	37.3003
210	2	0.4	210	4.5195	37.4373
210	2	0.4	270	4.5199	37.3709
310	2	0.4	210	4.5264	37.5909
110	2	0.4	210	4.5323	37.5726
510	2	0.4	210	4.5327	37.6709
310	2	0.5	270	4.5328	37.7425
410	2	0.4	210	4.5328	37.6765
110	2	0.5	270	4.5351	37.8490
810	2	0.4	210	4.5359	37.7415
910	2	0.4	210	4.5365	37.7524
710	2	0.4	210	4.5369	37.7752
210	2	0.5	270	4.5372	37.7365
610	2	0.4	210	4.5376	37.7699
610	2	0.5	270	4.5414	37.8141
810	2	0.5	210	4.5418	37.8655
710	2	0.5	270	4.5431	37.8322
410	2	0.5	210	4.5436	37.8748
610	2	0.5	210	4.5438	37.9102
710	2	0.5	210	4.5439	37.9067
410	2	0.5	270	4.5441	37.8995
910	2	0.5	210	4.5445	37.8973
510	2	0.5	210	4.5446	37.9445
510	2	0.5	270	4.5458	37.8803
810	2	0.5	270	4.5467	37.8629
910	2	0.5	270	4.5477	37.8813
110	2	0.4	270	4.5534	37.7081
310	2	0.5	210	4.5536	38.0264
110	2	0.5	210	4.5553	37.9823
210	2	0.5	210	4.5571	38.0741
910	2	0.4	150	4.5683	38.1897
510	2	0.6	210	4.5718	38.4476
410	2	0.6	210	4.5733	38.4577
810	2	0.4	150	4.5740	38.2876
310	2	0.4	150	4.5740	38.2130
710	2	0.6	210	4.5771	38.5153
810	2	0.6	210	4.5780	38.5060
610	2	0.6	210	4.5787	38.5224
910	2	0.6	210	4.5794	38.5356
710	2	0.4	150	4.5803	38.3878

n_estimators	min_samp_leaf	max_feat	trainDays	mae	mse
610	2	0.4	150	4.5825	38.4360
110	2	0.4	150	4.5826	38.3769
910	4	0.4	270	4.5846	38.4569
410	2	0.4	150	4.5847	38.4989
110	2	0.6	210	4.5855	38.6016
310	2	0.6	210	4.5857	38.6451
710	4	0.4	270	4.5858	38.4665
810	4	0.4	270	4.5866	38.4727
310	4	0.4	270	4.5876	38.3915
510	2	0.4	150	4.5879	38.5083
610	4	0.4	270	4.5890	38.5127
210	2	0.6	210	4.5891	38.6937
210	2	0.4	150	4.5893	38.4101
210	4	0.4	270	4.5901	38.3118
710	2	0.4	330	4.5902	38.6286
910	2	0.5	150	4.5909	38.6119
510	4	0.4	270	4.5920	38.5259
810	2	0.5	150	4.5925	38.6648
910	2	0.4	330	4.5926	38.6177
810	2	0.4	330	4.5935	38.6541
710	2	0.5	150	4.5939	38.6805
510	2	0.4	330	4.5941	38.6846
410	4	0.4	270	4.5945	38.5219
610	2	0.4	330	4.5947	38.6883
110	4	0.4	270	4.5975	38.4651
610	2	0.5	150	4.5993	38.7328
210	2	0.6	270	4.6008	38.9295
410	2	0.5	150	4.6011	38.7144
210	2	0.5	150	4.6031	38.6389
310	2	0.5	150	4.6033	38.6638
410	2	0.4	330	4.6034	38.7673
310	4	0.5	270	4.6043	38.7807
510	2	0.5	150	4.6052	38.8227
910	4	0.4	210	4.6056	38.7923
210	4	0.5	270	4.6058	38.7349
710	4	0.4	210	4.6062	38.7388
110	2	0.5	150	4.6064	38.5714
810	4	0.4	210	4.6066	38.7743
110	4	0.5	270	4.6066	38.8018
310	2	0.4	330	4.6078	38.8817
110	2	0.6	270	4.6080	38.9524
810	4	0.5	210	4.6090	38.9713
910	4	0.5	210	4.6093	38.9685
510	4	0.4	210	4.6116	38.7778
710	4	0.5	210	4.6117	39.0168
610	4	0.4	210	4.6122	38.8306

n_estimators	min_samp_leaf	max_feat	trainDays	mae	mse
310	2	0.6	270	4.6122	39.0505
910	2	0.6	270	4.6143	39.0849
610	4	0.5	210	4.6149	39.0805
710	2	0.6	270	4.6150	39.0574
810	2	0.6	270	4.6161	39.0965
410	4	0.5	210	4.6161	39.0368
910	4	0.5	270	4.6163	39.0564
610	2	0.6	270	4.6164	39.0691

Tabla A.3: Resultados hiperparametrización de Random Forest.

alpha	norm	mae	mse
1.0000	robust	5.3060	53.7996
0.0050	robust	5.3387	53.0297
0.1000	standard	5.3754	63.1635
0.5000	robust	5.3791	56.4442
0.0001	standard	5.3796	58.6247
0.0000	standard	5.3949	58.2504
0.0500	standard	5.3978	73.4872
0.0100	robust	5.4003	58.7508
0.0000	robust	5.4016	59.4235
1.0000	standard	5.4071	60.1640
0.0000	standard	5.4147	62.1670
0.0050	standard	5.4342	62.2698
0.0000	robust	5.4474	55.0361
0.5000	standard	5.4630	61.8776
0.0001	robust	5.4694	61.4467
0.0000	standard	5.4716	58.7974
0.0500	robust	5.4773	58.1999
0.0000	robust	5.4788	65.1034
0.1000	robust	5.4883	56.5177
0.0100	standard	5.4914	63.5145
0.0001	robust	5.4949	58.4795
0.0010	robust	5.4960	62.3716
0.0005	robust	5.5012	60.3506
0.0005	standard	5.5204	69.9221
0.0000	standard	5.5215	61.4272
0.0010	standard	5.5406	72.0959
0.0000	robust	5.5426	59.7882
0.5000	minMax	5.5495	59.9039
0.0000	minMax	5.5715	60.1123
0.0001	standard	5.5859	61.8129
0.0005	maxAbs	5.6312	56.5378
0.0500	minMax	5.6403	60.9515

alpha	norm	mae	mse
0.0500	maxAbs	5.6408	54.6134
0.0050	minMax	5.6520	56.7980
0.0050	maxAbs	5.6571	55.0113
0.1000	minMax	5.6606	59.3779
0.0010	maxAbs	5.6665	55.0562
0.0005	minMax	5.6714	58.9734
0.0100	maxAbs	5.6724	54.3429
0.5000	maxAbs	5.6741	57.1002
0.0000	maxAbs	5.6932	55.2853
0.0000	minMax	5.7067	58.0040
0.0000	maxAbs	5.7076	56.7831
0.0000	minMax	5.7152	60.1724
0.1000	maxAbs	5.7157	57.2769
1.0000	maxAbs	5.7175	57.5619
0.0001	minMax	5.7206	59.1069
0.0001	maxAbs	5.7278	59.1916
0.0000	maxAbs	5.7325	56.5400
0.0000	minMax	5.7352	57.6352
0.0000	maxAbs	5.7420	55.4517
0.0010	minMax	5.7665	62.2462
0.0001	maxAbs	5.7840	57.0907
0.0001	minMax	5.7844	63.8344
0.0100	minMax	5.8236	64.9056
1.0000	minMax	5.8316	60.3277

Tabla A.4: Resultados hiperparametrización de Perceptrón Multicapa.

C	epsilon	gamma	norm	trainDays	mae	mse
100	1	1	minMax	210	4.9312	45.2712
100	1	0.0100	standard	150	4.9693	44.2717
100	1	1	maxAbs	210	4.9705	44.9035
100	0.1000	0.0100	standard	150	4.9794	44.5224
100	0.0100	0.0100	standard	150	4.9821	44.5327
100	0.0010	0.0100	standard	150	4.9826	44.5383
100	0.0001	0.0100	standard	150	4.9827	44.5389
100	0	0.0100	standard	150	4.9827	44.5391
100	0.1000	1	minMax	210	4.9948	46.0844
100	1	1	minMax	270	4.9971	43.8522
100	1	0.1000	robust	210	5.0009	44.5691
100	0.0100	1	minMax	210	5.0045	46.2223
100	0.0010	1	minMax	210	5.0052	46.2300
100	0.0001	1	minMax	210	5.0052	46.2302
100	0	1	minMax	210	5.0053	46.2306
100	1	1	maxAbs	270	5.0073	43.9694

C	epsilon	gamma	norm	trainDays	mae	mse
100	1	0.1000	robust	270	5.0218	44.2292
10	0.0001	0.1000	robust	210	5.0222	44.8433
10	0	0.1000	robust	210	5.0222	44.8435
10	0.0010	0.1000	robust	210	5.0223	44.8438
10	0.0100	0.1000	robust	210	5.0225	44.8475
100	0.1000	1	maxAbs	210	5.0246	45.7534
10	1	0.1000	robust	210	5.0247	44.7108
10	0.1000	0.1000	robust	210	5.0254	44.8884
100	0.0100	1	maxAbs	210	5.0265	45.7616
100	0	1	maxAbs	210	5.0265	45.7687
100	0.0001	1	maxAbs	210	5.0265	45.7687
100	0.0010	1	maxAbs	210	5.0266	45.7684
100	0.1000	1	minMax	270	5.0314	44.4253
10	1	0.1000	robust	150	5.0327	46.0169
100	0.0100	1	minMax	270	5.0395	44.5365
100	0.0100	0.1000	robust	210	5.0402	45.3881
100	0.0010	1	minMax	270	5.0402	44.5471
100	0.0010	0.1000	robust	210	5.0402	45.3949
100	0	0.1000	robust	210	5.0402	45.3958
100	0.0001	0.1000	robust	210	5.0402	45.3958
100	0	1	minMax	270	5.0403	44.5476
100	0.0001	1	minMax	270	5.0403	44.5477
100	0.1000	0.1000	robust	210	5.0419	45.3692
10	0.1000	0.1000	robust	150	5.0479	46.2779
100	1	0.0100	standard	90	5.0486	45.4846
10	0.0100	0.1000	robust	150	5.0497	46.3087
10	0.0010	0.1000	robust	150	5.0498	46.3076
10	0.0001	0.1000	robust	150	5.0498	46.3075
10	0	0.1000	robust	150	5.0498	46.3076
10	0.1000	0.0100	standard	90	5.0627	46.3395
10	1	0.0100	standard	90	5.0638	46.2423
10	0.0100	0.0100	standard	90	5.0648	46.3469
10	0.0001	0.0100	standard	90	5.0649	46.3494
10	0	0.0100	standard	90	5.0649	46.3494
10	0.0010	0.0100	standard	90	5.0649	46.3493
100	0.0100	0.1000	robust	270	5.0725	45.0639
100	0.1000	0.1000	robust	270	5.0728	45.0670
100	0.0010	0.1000	robust	270	5.0731	45.0745
100	0.0001	0.1000	robust	270	5.0732	45.0759
100	0	0.1000	robust	270	5.0732	45.0762
100	0.1000	1	maxAbs	270	5.0748	45.0689
100	0.0100	1	maxAbs	270	5.0750	45.0922
100	0.0010	1	maxAbs	270	5.0752	45.0968
100	0	1	maxAbs	270	5.0752	45.0970
100	0.0001	1	maxAbs	270	5.0753	45.0971
10	1	0.1000	robust	270	5.0794	44.6692

C	epsilon	gamma	norm	trainDays	mae	mse
10	1	1	minMax	210	5.0801	46.4409
100	1	0.1000	standard	210	5.0823	47.9746
100	0.1000	0.0100	standard	90	5.0851	46.3185
100	0.0100	0.0100	standard	90	5.0871	46.3403
100	0.0010	0.0100	standard	90	5.0875	46.3462
100	0.0001	0.0100	standard	90	5.0875	46.3466
100	0	0.0100	standard	90	5.0875	46.3467
100	1	0.1000	standard	270	5.0890	46.9038
100	1	0.0100	robust	150	5.0905	45.5480
100	1	0.0100	standard	210	5.0912	44.9042
10	0	0.1000	robust	270	5.0970	45.1369
10	0.0001	0.1000	robust	270	5.0970	45.1368
10	0.0010	0.1000	robust	270	5.0970	45.1365
10	0.0100	0.1000	robust	270	5.0971	45.1363
10	0.1000	0.1000	robust	270	5.0986	45.1312
100	1	0.1000	robust	150	5.1009	47.6777
10	0.1000	1	minMax	210	5.1108	47.0976
10	0.0001	1	minMax	210	5.1119	47.1236
10	0.0010	1	minMax	210	5.1119	47.1234
10	0	1	minMax	210	5.1119	47.1236
10	0.0100	1	minMax	210	5.1119	47.1219
100	1	0.0100	robust	90	5.1120	47.2203
100	0	0.0100	standard	210	5.1153	45.3491
100	0.0001	0.0100	standard	210	5.1154	45.3492
100	0.0010	0.0100	standard	210	5.1154	45.3500
100	0.0100	0.0100	standard	210	5.1160	45.3615
100	0.1000	0.0100	robust	150	5.1166	46.0750
100	0.0100	0.0100	robust	150	5.1174	46.0979
100	0.0010	0.0100	robust	150	5.1174	46.1013
100	0	0.0100	robust	150	5.1175	46.1022
100	0.0001	0.0100	robust	150	5.1175	46.1025
100	0.1000	0.0100	standard	210	5.1201	45.4221
100	0.1000	0.1000	standard	210	5.1296	48.7471
100	0.1000	0.0100	robust	90	5.1343	47.6832
100	0.0100	0.0100	robust	90	5.1355	47.6860
100	0.0001	0.0100	robust	90	5.1355	47.6851
100	0.0010	0.0100	robust	90	5.1355	47.6854
100	0	0.0100	robust	90	5.1355	47.6852
1000	1	0.0100	standard	210	5.1366	47.0656

Tabla A.5: Resultados hiperparametrización de los modelos SVM.